UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMÁTICAS
Departamento de Estadistica e Investigacion Operativa

TESIS DOCTORAL

Problemas de decisión Bayesiana no paramétrica : aproximación lineal

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Alfonso García Pérez

Madrid, 2015

Alfonso García Pérez

PROBLEMAS DE DECISION BAYESIANA NO PARAMETRICA: APROXIMACION LINEAL

Departamento de Estadística e Investigación Operativa
Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid
1983

Colección Tesis Doctorales. Nº 12/83

Edita e imprime la Editorial de la Universidad Complutense de Madrid. Servicio de Reprografía
Noviciado, 3 Madrid-8
Madrid, 1983
Xerox 9200 XB 480
Depósito Legal: M-1170-1983

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE MATEMATICAS

SECCION DE ESTADISTICA E I.O.

"PROBLEMAS DE DECISION BAYESIANA NO PARAMETRICA:

APROXIMACION LINEAL"

por

Alfonso García Pérez

Memoria realizada bajo la
dirección del Profesor Dr.
D. Vicente Quesada Paloma
para optar al grado de Doc
tor en Ciencias Matemáticas.

A mis padres

INDICE

INTRODUCCION

La introducción de una medida de probabilidad sobre el espacio paramétrico y la utilización del teorema de Bayes como base para la toma de decisiones o simplemente para la inferencia estadística se ha movido entre la aceptación y el rechazo desde la publicación de dicho Teorema en 1763.

Podemos afirmar que en la situación paramétrica en donde se conoce la forma de las distribuciones y en donde la estimación de ciertos caracteres desconocidos de ellas llamados parámetros es el objetivo, el Bayesianismo ha tenido éxito. No obstante, cuando ni siquiera la forma de las distribuciones es conocida, es decir, cuando es la propia distribución el parámetro a estudiar y estimar, y el espacio paramétrico el conjunto de todas las distribuciones de probabilidad definidas en un espacio muestral dado, el Bayesianismo, la teoría de la decisión Bayesiana no paramétrica ha encontrado muy serias dificultades y su aplicación no ha sido muy amplia.

Ferguson en 1973, utilizando los procesos de Dirichlet como inductores de medidas de probabilidad sobre el espacio paramétrico y más tarde Doksum en 1974, empleando éste los procesos neutrales por la derecha como inductores de medidas de probabilidad marcan la pauta de la decisión Bayesiana no paramétrica, aunque los problemas en la determinación de la distribución a posteriori y en la esperanza respecto de ésta hacen difícil y laboriosa la resolución de problemas concretos de estimación.

Goldstein en 1975 propone un método alternativo: restringir el conjunto de reglas de decisión en la búsqueda de la regla Bayes, utilizando métodos lineales.

Esta idea fue punto de partida en la realización de la presente tesis doctoral. Nuestro estudio se basará en la restricción de la búsqueda de reglas de decisión en el conjunto de combinaciones lineales de algún conjunto dado de funciones de la muestra.

El presente trabajo estudiará variables aleatorias las cuales representan tiempo transcurrido hasta que un evento ocurre. Tales eventos, o sucesos en la terminología habitual, son referidos como "fallos", aunque éstos pueden ser por ejemplo la realización de un cierto trabajo en un experimento de aprendizaje en psicología, o el cambio de residencia en un estudio demográfico, o simplemente la avería de un mecanismo inanimado. No obstante, su mayor utilidad reside en los estudios médicos tales como la investigación del cáncer.

Así pues, lo que haremos será estudiar variables aleatorias $T \geq 0$, que representen tiempos de fallo de individuos de una población homogénea, comenzando con un estudio probabilístico profundo de las funciones de supervivencia aleatorias en el primer capítulo, en donde estableceremos los principales teoremas, los cuales serán las herramientas que utilizaremos en las estimaciones del resto del trabajo. Supondremos tres diferentes distribuciones a priori: la inducida por un proceso gamma exponencial en el segundo capítulo, la inducida por un proceso homogéneo simple en el tercero y la inducida por un proceso gamma extendido en el cuarto, siendo los dos primeros procesos neutrales a la derecha y no siendolo el proceso gamma extendido, introducido por Dykstra y Purushottam en 1981, y utilizando a cada una de ellas como distribución a priori determinaremos estimadores de la función de supervivencia, la función de distribución, el tiempo medio de supervivencia, el momento de orden i, trataremos el problema de las dos muestras, etc.

Supondremos también situaciones en las cuales alguno o algunos de nuestros tiempos de fallo en estudio sean "censurados", es decir, consideraremos situaciones en las cuales en alguno de nuestro individuos en estudio se produzca el fallo por una causa distinta a la que estamos estudiando. En éstas situaciones, una idea errónea sería el no considerar a dichos individuos, ya que ellos nos proporcionan la información de que su tiempo de fallo es mayor al acaecido por esa causa extraña, evitando por otro lado la subjetividad en la eliminación de muestras seleccionadas.

Hay que notar que también Ferguson y Phadia (1979) han determinado estimaciones de la función de supervivencia, tanto en el caso de que haya datos censurados como en el caso de que no los haya, mediante el cálculo de la media a posteriori. Sin embargo, las estimaciones por ellos efectuadas resultan totalmente inmanejables, aún en el caso de no considerar datos censurados, y los parámetros que en sus estimaciones intervienen no presentan ninguna significación clara. Nuestras estimaciones, basadas, como ya hemos dicho, en la linealidad de las reglas Bayes consideradas, nos conducirán a estimaciones totalmente útiles y aplicables a cualquier tipo de situaciones, inclusive al caso en el que se presenten datos censurados, y con una significación muy clara de los parámetros que en ellas intervienen.

Estudiaremos también los comportamientos asintóticos de todo tipo, y observaremos en ellos una clara conducta Bayesiana.

Sin embargo, no quedaría completo éste trabajo sin una comparación con el de Ferguson y Phadia para ver cual ha sido la pérdida de calidad por el aumento de utilidad. Así lo haremos, y los resultados que se observarán serán sorprendentes: Podremos afirmar que el estimador de la función de supervivencia propuesto por Ferguson y Phadia actúa como curva de regresión, mientras que el nuestro lo hace como recta de regresión.

Así pues, los objetivos de éste trabajo son múltiples, aunque pueden dividirse en dos grandes bloques: en el primero, que comprende el primer capítulo, haremos los estudios probabilísticos y de procesos que trae consigo la utilización de funciones de supervivencia, distribución, etc., aleatorias. En el segundo, que comprende los capítulos segundo, tercero y cuarto, nos centraremos en un estudio estadístico calculando reglas Bayes y sus comportamientos asintóticos, para diversos parámetros, uno de ellos la propia función de supervivencia, así como sus riesgos Bayes asociados.

Específicamente, en el capítulo primero empezaremos definiendo en la sección I el marco en el cual nos vamos a mover, cual es el de la teoría de la decisión Bayesiana no paramétrica. Continuaremos en la sección II definiendo las principales funciones que vamos a manejar, como son la función de supervivencia, la función tasa de azar, la función tasa de azar acumulativa, así como la función de supervivencia muestral. Es la sección III una sección fundamental, ya que en ella consideraremos el concepto de función aleatoria, así como algunas propiedades, tanto probabilísticas como estadísticas, de algunas de dichas funciones aleatorias en el análisis de la supervivencia. Algunos de dichos resultados, todos ellos totalmente originales, son por ejemplo el dado en el teorema 1.3.4, el cual nos dice que bajo ciertas condiciones muy generales, dada una variable aleatoria T con función de supervivencia aleatoria $S(t)$, siendo g una función medible de T y siendo P una distribución de probabilidad a priori sobre el espacio paramétrico F', es

$$\int_{F'} \left(\int_0^\infty g(t)S(t)dt\right)\left(\int_0^\infty g(u)S(u)du\right)dP(S) =$$
$$= \int_0^\infty \int_0^\infty g(t)g(u) \left(\int_{F'} S(t)S(u)dP(S)\right)dt\ du$$

La posibilidad de operar con dos funciones g_1 y g_2 en vez de sólo con una, nos la da el teorema 1.3.5, en donde ahora

$$\int_{F'} \left(\int_0^\infty g_1(t)S(t)dt\right) \left(\int_0^\infty g_2(u)S(u)du\right)dP(S) =$$
$$= \int_0^\infty\int_0^\infty g_1(t)g_2(u) \left(\int_{F'} S(t)S(u)dP(S)\right)dt\ du.$$

También el teorema 1.3.6 nos va a dar el resultado de ser

$$\int_{F'} \left(\int_0^\infty S(t)g(t)dt\right)dP(S) = \int_0^\infty g(t)\left(\int_{F'} S(t)dP(S)\right)dt$$

No obstante, si $F(t) = 1 - S(t)$ es la correspondiente función de distribución aleatoria, nuestro interés fundamental será el determinar

$$\int_F \left(\int_0^\infty h_1(x)dF(x)\right)\left(\int_0^\infty h_2(y)dF(y)\right)d\mathcal{P}(F)$$

en donde h_1 y h_2 son dos funciones medibles de T; pues bien, el teorema 1.3.7 nos da la solución, siendo dicha integral igual a

$$c_1 \cdot c_2 - c_1 \int_0^\infty g_2(y)\left(\int_{F'} S(y)d\mathcal{P}(S)\right)dy -$$

$$- c_2 \int_0^\infty g_1(x)\left(\int_{F'} S(x)d\mathcal{P}(S)\right)dx +$$

$$+ \int_0^\infty\int_0^\infty g_1(x)g_2(y)\left(\int_{F'} S(x)S(y)d\mathcal{P}(S)\right)dx\, dy,$$

en donde c_1 y c_2 son dos constantes allí especificadas y $g_1(x)$ y $g_2(y)$ son respectivamente las derivadas respecto a la medida de Lebesgue de $h_1(x)$ y $h_2(y)$.

En todos éstos resultados, ya digo totalmente originales, hemos dicho que teníamos la solución cuando las integrales se transformaban en otras donde aparecían integrales del tipo

$$\int_{F'} S(t)d\mathcal{P}(S) = E[S(t)]$$

o integrales del tipo,

$$\int_{F'} S(x)S(y)d\mathcal{P}(S) = E[S(x)\; S(y)]$$

pero, ¿es que acaso éstas últimas son más fáciles de determinar?; la respuesta es si, y de hecho éste es uno de los objetivos de las secciones VI, VII y VIII de éste capítulo primero, cuando $\mathcal{P}$ es la inducida por un proceso gamma exponencial, homogéneo simple y gamma extendido respectivamente, procesos éstos últimos introducidos éste año y que no son neutrales a la derecha, a diferencia de los gamma exponenciales y homogéneo simples que si lo son, obteniendose respuestas totalmente originales a dichos objetivos.

Hemos dicho que en la sección III antes mencionada obteníamos también propiedades estadísticas. ¿Cuáles son éstas?. Sin duda, la fundamental es la estimación de una función de supervivencia $S(t)$, expresada a lo largo de los teoremas 1.3.8, 1.3.9 y 1.3.10, así como en sus corolarios asociados, resultados todos ellos originales, y que resumidos vienen a decir que si consideramos una función de supervivencia aleatoria $S(t)$ y una medida de probabilidad aleatoria P a priori, sobre el espacio paramétrico F', consideramos una muestra de tamaño n_1 y buscamos la regla Bayes $\hat{S}_1(t)$ para $S(t)$ dentro de la clase de reglas de decisión de la forma

$$a\, S_{n_1}(t) + b\, E[S(t)]$$

a continuación extraemos una muestra de tamaño n_2 y buscamos la regla Bayes, ahora dentro de la clase de reglas de decisión de la forma

$$a\, S_{n_2}(t) + b\, \hat{S}_1(t),$$

continuamos el proceso, siendo n_k el tamaño de la muestra en la k-ésima y última etapa y buscando en dicha última etapa dentro de la clase de reglas de decisión de la forma

$$a\, S_{n_k}(t) + b\, \hat{S}_{k-1}(t),$$

la regla Bayes es

$$\hat{S}_k(t) = [1 - p_{n_k}(t)]\, S_{n_k}(t) + p_{n_k}(t)\, E[S(t)]$$

"

siendo el riesgo Bayes el dado en el mencionado teorema 1.3.9, y siendo $S_{n_i}(t)$ la función de supervivencia muestral,

$$p_{n_k}(t) = \frac{E[S(t)] - E[S^2(t)]}{E[S(t)] + (n_k-1)\ E[S^2(t)] - n_k\ (E[S(t)])^2}$$

y estando todas las esperanzas calculadas con respecto a P. Vemos allí también que $\hat{S}_k(t)$ o $\hat{S}(t)$ en general por abreviar, converge casi seguro hacia $S(t)$ y más tarde, cuando estudiemos distribuciones a priori P particulares, la inducida por un proceso gamma exponencial, por un proceso homogéneo simple y por uno gamma extendido, intervendrán unos parámetros, en particular uno, c, el cual nos va a medir el grado de confianza de la estimación a priori $S_o(t) = E[S(t)]$ y que cuando $c \to o$ ($S_o(t)$ estimación a priori pésima), $\hat{S}(t)$ sólo dependerá de la información muestral convirtiendose $\hat{S}(t) = S_{n_k}(t)$, siendo $\hat{S}(t) = S_o(t)$ cuando $c \to \infty$ ($S_o(t)$ estimación a priori óptima). Vemos también en ésta sección III que ocurre al habernos restringido a clases de reglas de decisión lineales en vez de haber considerado el usual estimador de la media a posteriori, ya que siempre consideraremos funciones de pérdida cuadráticas. La respuesta es que la media a posteriori actúa como la curva de regresión de $S(t)$ sobre $S_n(t)$ y nuestro estimador como la correspondiente recta de regresión.

Sin darnos cuenta, después de hacer un resumen de todos estos resultados totalmente originales y que luego demostraremos, hemos entrado en los capítulos segundo, tercero y cuarto, los cuales son totalmente originales. En ellos, aunque el esquema de trabajo es el mismo en los tres, la distribución a priori es totalmente diferente: en el segundo capítulo supondremos que $\mathcal{P}$ es la inducida por un proceso gamma exponencial, en el tercero $\mathcal{P}$ será la inducida por un proceso homogéneo simple y en el cuarto $\mathcal{P}$ lo será por un proceso gamma extendido, con lo que las conclusiones e interpretaciones de dichos resultados serán diferentes. En los tres empezaremos con el tratamiento antes mencionado en la estimación de la función de supervivencia, primero en el caso de que no haya datos censurados y luego en el supuesto de que los haya. Las siguientes secciones se dedican a la estimación de la función de distribución, el tiempo medio de supervivencia, el momento de orden i, el tratamiento del problema de las dos muestras, etc., siguiendo un procedimiento semejante al seguido en la estimación de la función de supervivencia, y obteniendo resultados paralelos a los allí obtenidos, especialmente en cuanto a la interpretación de los parámetros.

Como se ve la estimación de la función de supervivencia, en el segundo bloque de éste trabajo, dedicado a estimaciones, tiene una importancia decisiva y fundamental. ¿Por qué es esto?. La respuesta ahora y con ésto termino, nos la da los teoremas 2.1.2 y 2.1.3, sin menospreciar el resultado alcanzado en el 2.1.1, en donde demostramos, ya que también dichos teoremas son originales, que si $\hat{S}(t)$ es la regla Bayes, dentro de alguna clase de reglas de decisión, para $S(t)$, y si g es una función lineal (bilineal) $g(\hat{S}(t))$ es la regla Bayes para $g(\hat{S}(t))$ dentro de la clase de reglas de decisión imágenes mediante g de la clase en donde $\hat{S}(t)$ era regla Bayes para $S(t)$, manteniendose los mismos riesgos en dichas transformaciones, ya que por supuesto, en todas las estimaciones damos los riesgos Bayes mínimos alcanzados.

No quiero terminar sin agradecer a mi profesor y amigo el Dr. D. Vicente Quesada Paloma, bajo cuya dirección ha sido realizado éste trabajo, no solo el haber hecho posible la realización del mismo con sus orientaciones, enseñanzas y discusiones, sino también el haberme permitido trabajar a su lado durante los tres últimos años. A mi amigo el Dr. D. Michael Goldstein, todas las sugerencias y discusiones que me brindó durante dos meses de trabajo intenso en Hull (Inglaterra), y a todos mis compañeros de Departamento y en especial a su jefe, los cuales siguieron con interés las incidencias de éste trabajo y con quien discutí algunos puntos del mismo.

Madrid, Diciembre, 1981

CAPITULO 1

ELEMENTOS BASICOS

I. DECISION BAYESIANA NO PARAMETRICA

En todo problema de decisión se pueden encontrar tres elementos básicos: Un espacio de estados posibles de la Naturaleza Θ, llamado también espacio paramétrico y que contiene a la cantidad desconocida o parámetro θ sobre el que se quiere decidir o investigar; un espacio de acciones posibles a tomar por el decisor A, y una función de pérdida $L : \Theta \times A \longrightarrow R$.

En un problema de decisión no paramétrico, el espacio paramétrico es el espacio de todas las distribuciones de probabilidad definidas en un espacio muestral dado; se representa por F.

La utilización del procedimiento Bayesiano como criterio de decisión lleva implícito la definición de una distribución de probabilidad P sobre el espacio paramétrico, llamada distribución de probabilidad a priori.

Con objeto de obtener información acerca de θ, se realiza una investigación estadística cuyo resultado es una variable aleatoria que será denotada por X, pudiendo representar también X a un vector aleatorio. El conjunto de posibles resultados es el espacio muestral que será denotado por $\mathcal{X}$. Llamaremos regla de decisión $d(x)$ a una función medible de $\mathcal{X}$ en A, supuesto definidas dos σ-álgebras convenientes en A y $\mathcal{X}$.

La variable aleatoria X tendrá una distribución que dependerá del parámetro. Llamaremos función de riesgo de una regla de decisión $d(x)$ a la esperanza, con respecto a la distribución de X, de la pérdida $L(\theta,\delta(X))$. La representaremos por $R(\theta,d)$.

Obsérvese que una acción de A es un caso particular de regla de decisión cuando no se realiza experimento alguno, coincidiendo en dicho caso la pérdida y el riesgo.

Definición 1.1.1. Llamaremos riesgo Bayes de una regla de decisión d, correspondiente a una distribución a priori P, y lo representaremos por $r(P,d)$, a la esperanza con respecto a P de la función de riesgo de la regla $d(x)$; es decir,

$$r(P,d) = E_P\left[R(\theta,d)\right]$$

Definición 1.1.2. Llamaremos regla Bayes con respecto a una distribución a priori P a la regla de decisión que haga mínimo el riesgo Bayes correspondiente a P. Analíticamente,

$$d^* \text{ regla Bayes correspondiente a } P \overset{def}{\Longleftrightarrow} r(P,d^*) = \min_d r(P,d)$$

Definición 1.1.3. Llamaremos riesgo Bayes de P a la cantidad

$$r(P) = r(P,d^*)$$

en donde d^* es la regla Bayes con respecto a P.

Definición 1.1.4. Si A es el espacio de acciones antes considerado, una regla de decisión aleatorizada $\delta^*(x,.)$ es para cada x, una distribución de probabilidad sobre A.

Definición 1.1.5. La función de pérdida $L(\theta,\delta^*(x,.))$ de una regla de decisión aleatorizada δ^* se define como

$$L(\theta,\delta^*(x,.)) = E_{\delta^*(x,.)}\left[L(\theta,a)\right]$$

definiendose la función de riesgo por tanto como

$$R(\theta,\delta^*) = E_X\left[L(\theta,\delta^*(X,.))\right]$$

Conviene aclarar algo en cuanto a la notación: cuando calculemos esperanzas con respecto a una distribución de probabilidad π, lo representaremos por $E_\pi[.]$, representando por $E_Y[.]$ la esperanza calculada con respecto a la distribución de la variable aleatoria Y.

<u>Definición 1.1.6</u>. Sea $\mathcal{D}^*$ el conjunto de reglas de decisión aleatorizadas δ^* para las cuales $R(\theta,\delta^*) < \infty \quad \forall\theta \in \Theta$. Diremos que una regla δ_1^* es R-mejor que otra δ_2^* si $R(\theta,\delta_1^*) \leq R(\theta,\delta_2^*) \quad \forall\theta \in \Theta$ con la igualdad estricta en algún $\theta \in \Theta$.

Diremos que una regla δ_1^* es equivalente a otra δ_2^* si $R(\theta,\delta_1^*) = R(\theta,\delta_2^*) \quad \forall\theta \in \Theta$.

Conviene notar que las reglas de decisión no aleatorizadas serán consideradas como un caso particular de regla de decisión aleatorizada y consecuentemente las acciones.

<u>Teorema 1.1.1</u>. (Berger (1.980)): Sea T un estadístico suficiente para $\theta \in \Theta$, y sea $\delta_o^*(x,.)$ una regla de decisión aleatorizada de $\mathcal{D}^*$. Entonces, la regla de decisión aleatorizada

$$\delta_1^*(t,.) = E_{X/t}\left[\delta_o^*(X,.)\right]$$

la cual depende solamente de $T(x)$, es R-equivalente a $\delta_o^*(x,.)$.

II. LA FUNCION DE SUPERVIVENCIA

Como antes mencionamos, estudiaremos variables aleatorias las cuales representan tiempos de fallo, es decir tiempo transcurrido hasta que un determinado suceso ocurre.

Definición 1.2.1. Sea $T \geq 0$ una variable aleatoria no negativa definida sobre el espacio probabilístico $(\Omega, \sigma_\Omega, P)$, la cual representa el tiempo de muerte de un individuo de una población homogénea, o tiempo de fallo de un sistema. Llamaremos función de supervivencia y la representaremos por $S(t)$ a

$$S(t) = P\{T > t\}$$

Proposición 1.2.1:

$S(t)$ es una función de supervivencia si y sólo si

(1) $S(t)$ es monótona no creciente

(2) $S(t)$ es contínua a la derecha de cada punto.

(3) $\lim\limits_{t \to o} S(t) = 1$

(4) $\lim\limits_{t \to \infty} S(t) = 0$

La demostración es evidente teniendo en cuenta que $S(t) = 1-F(t)$, en donde $F(t)$ es la función de distribución asociada a T, y por las propiedades que caracterizan a las funciones de distribución.

Asociada a T existe otra función muy utilizada y que es llamada función tasar de azar.

Definición 1.2.2. Sea $T \geq 0$ la variable aleatoria antes considerada. Si T es discreta tomando los valores $x_1 < x_2 < \dots,$ llamaremos función tasa de azar correspondiente a T, a la función

$$\lambda_j = \frac{p(x_j)}{S(x_j)} , \qquad j=1,2,\dots$$

en donde $p(x_j) = P\{T = x_j\}$, $j=1,2,\dots,$ es la función de masa asociada y $S(x_j)$ la de supervivencia.

Si T es continua, llamaremos función de tasa de azar, a la función

$$\lambda(t) = \frac{f(t)}{S(t)}$$

siendo $f(t)$ la función de densidad asociada a T.

Como se ve, la función $\lambda(t)$ indica la propensión de fallo de un "item" o individuo de la población considerada, en un futuro cercano dado que el individuo ha sobrevivido hasta el tiempo t.

Definición 1.2.3. Sea $T \geq 0$ la variable aleatoria antes considerada. Si T es discreta tomando los valores $x_1 < x_2 < \dots,$ llamaremos función de azar acumulativa correspondiente a T, a la función

$$H_j = \sum_{i=1}^{j} \lambda_i , \qquad j=1,2,\dots$$

en donde λ_i es la tasa de azar correspondiente.

Si T es continua, llamaremos función de azar acumulativa a la función,

$$H(t) = \int_0^t \lambda(s)ds = -\log S(t)$$

en donde $\lambda(s)$ es la correspondiente función tasa de azar y log indica logaritmo neperiano.

Definición 1.2.4. Sea $T_1,\ldots,T_n$ una muestra aleatoria simple de la variable aleatoria T antes considerada. Llamaremos función de supervivencia muestral, y la representaremos por $S_n(t)$, a la variable aleatoria

$$S_n(t) = \frac{n^\circ \text{ de } T_i > t}{n}$$

Proposición 1.2.2. La función de supervivencia muestral verifica que

(a) $E[S_n(t)] = S(t)$

(b) $V(S_n(t)) = \frac{S(t) \cdot (1 - S(t))}{n}$

(c) $E[S_n^2(t)] = \frac{S(t)}{n} + \frac{n-1}{n} S^2(t)$

En efecto:

Si llamamos X a la variable aleatoria "número de $T_i > t$", X seguirá una distribución binomial $B(n, S(t))$, por ser $p = Pr\{T_i > t\} = Pr\{T > t\} = S(t)$. Así pues,

$$E[X] = n.S(t)$$

$$V(X) = n.S(t).(1 - S(t))$$

con lo que

$$E[S_n(t)] = E\left[\frac{X}{n}\right] = S(t)$$

y

$$V(S_n(t)) = \frac{1}{n^2} V(X) = \frac{S(t)\cdot(1 - S(t))}{n}$$

siendo por tanto,

$$E\left[S_n^2(t)\right] = V(S_n(t)) + (E\,[S_n(t)])^2 = \frac{S(t)}{n} + \frac{n-1}{n}\,S^2(t)$$

como queríamos demostrar.

<u>Teorema 1.2.1</u>. La función de supervivencia muestral $S_n(t)$ es un estimador suficiente para $S(t)$ para una muestra aleatoria simple de tamaño n.

En efecto:

Sea $T_1,\ldots,T_n$ una muestra aleatoria simple de la variable aleatoria T. Definimos las variables aleatorias

$$Y_i = \begin{cases} 1 & \text{si} \quad T_i > t \\ 0 & \text{si} \quad T_i \leq t \end{cases}$$

Como las T_i son independientes, entonces las Y_i también lo son. Además, como $Pr\{Y_i = 1\} = Pr\{T_i > t\} = S(t)$, serán

$$Y_i \equiv B(1,S(t))$$

Consideremos el estimador $T = T(Y_1,\ldots,Y_n) = \sum_{i=1}^{n} Y_i$. Entonces $T \equiv B(n,S(t))$ y

$$Pr\{Y_1 = y_1,\ldots,Y_n = y_n \,/\, \sum_{i=1}^{n} Y_i = t'\} =$$

$$= \frac{Pr\{Y_i = y_1,\ldots,Y_n = y_n,\ T = t'\}}{\binom{n}{t'}[S(t)]^{t'}\,[1 - S(t)]^{n-t'}}$$

si $\sum_{i=1}^{n} y_i = t'$ y <u>o</u> en otro caso.

Así pues, si $\sum_{i=1}^{n} y_i = t'$

$$Pr\{Y_1 = y_1,\ldots,Y_n = y_n/T = t'\} =$$

$$= \frac{\Pr\{Y_1 = y_1\} \ \dots \ \Pr\{Y_n = y_n\}}{\binom{n}{t'} [S(t)]^{t'} [1-S(t)]^{n-t'}} =$$

$$= \frac{[S(t)]^{y_1} [1-S(t)]^{1-y_1} \dots [S(t)]^{y_n} [1-S(t)]^{1-y_n}}{\binom{n}{t'} [S(t)]^{t'} [1-S(t)]^{n-t'}}$$

Así pues,

$$\Pr\{Y_1 = y_1,\dots,Y_n = y_n/T=t'\} = \begin{cases} \dfrac{1}{\binom{n}{t'}} & \text{si} \quad \sum_{i=1}^{n} y_i = t' \\ 0 & \text{en el resto} \end{cases}$$

la cual no depende del parámetro $S(t)$. Obsérvese además que

$$T = \sum_{i=1}^{n} Y_i = n^\circ \text{ de } T_i > t = n\, S_n(t).$$

Así pues, $n\, S_n(t)$ es un estimador suficiente para $S(t)$, con lo que $S_n(t) = \frac{n \cdot S_n(t)}{n}$ es también un estimador suficiente para $S(t)$, por ser $g(x) = \frac{x}{n}$ una función estrictamente monótona y derivable (más concretamente una función uno a uno).

III. FUNCIONES ALEATORIAS EN EL ANALISIS DE SUPERVIVENCIA

Sea X un conjunto y A una σ-álgebra de subconjuntos de X. Sea pues el espacio medible (X,A).

Sea $P(\omega,A)$ un proceso estocástico con conjunto de índices A, con espacio de estados el intervalo $[0,1]$ y definido sobre un espacio probabilístico $(\Omega,\sigma_\Omega,\lambda)$.

Si fijamos un índice, es decir, si fijamos un $A \in A$,

$$P(A) = P(.,A) : \Omega \longrightarrow [0,1]$$

es una variable aleatoria.

Si fijamos un $\omega \in \Omega$,

$$P(\omega,.) : \mathcal{A} \longrightarrow [0,1]$$

es una trayectoria del proceso. Dicha trayectoria es, o mejor dicho, puede llegar a ser una medida de probabilidad, pero fijado un $A \in \mathcal{A}$, $P(A)$ es además de la probabilidad de A, una variable aleatoria. Llamaremos por tanto a P, medida de probabilidad aleatoria sobre $(X,\mathcal{A})$, o simplemente probabilidad aleatoria. Tenemos pues la siguiente definición.

<u>Definición 1.3.1</u>. Dado un conjunto X y una σ-álgebra $\mathcal{A}$ de subconjuntos de X, llamaremos medida de probabilidad aleatoria sobre el espacio medible $(X,\mathcal{A})$ a un proceso $\{P(A) : A \in \mathcal{A}\}$, definido sobre algún espacio probabilístico $(\Omega,\sigma_\Omega,\lambda)$ y con espacio de estados el intervalo $[0,1]$, tal que

(1) $P(A)$ es una variable aleatoria con valores en $[0,1]$ para $A \in \mathcal{A}$.

(2) $P(X) = 1$ c.s.

(3) $\lim\limits_{k\to\infty} P(A_k) = 0$ c.s. para cada sucesión decreciente $\{A_k\} \subset \mathcal{A}$ con $\lim\limits_{k\to\infty} A_k = \Phi$.

Como nosotros sólo trabajaremos en el caso real, de ahora en adelante se supondrá que $(X,\mathcal{A})$ es (R,B), el espacio medible de la recta real con la σ-álgebra de Borel.

Si hacemos uso de las propiedades de orden de la recta real, es natural introducir lo que llamaremos función de distribución aleatoria F, correspondiente a la probabilidad aleatoria P, que vendrá definida por

$$F(t) = P((-\infty,t]) \qquad (1)$$

El proceso estocástico así definido tiene una versión separable (un proceso con la misma distribución de probabilidad que F) el cual posee unas propiedades que permite caracterizarlo. Definiendo la función de distribución aleatoria por medio de éstas propiedades, como a continuación lo vamos a hacer, tendremos la ventaja de no depender, al estudiar una función de distribución aleatoria, de la probabilidad aleatoria asociada.

<u>Definición 1.3.2</u>. Llamaremos función de distribución aleatoria a un proceso estocástico $\{F(t) : t \in \mathbb{R}\}$, definido sobre algún espacio probabilístico $(\Omega,\sigma_\Omega,\lambda)$ y con espacio de estados el intervalo $[0,1]$, tal que

(1) $F(t)$ es monótona no decreciente c.s.

(1) $\lim\limits_{t\to-\infty} F(t) = 0$ c.s.

(3) $\lim\limits_{t\to+\infty} F(t) = 1$ c.s.

(4) $F(t)$ es contínua por la derecha casi seguramente, es decir,

$$\forall t \in \mathbb{R}, \quad \lim_{s\to t^+} F(s) = F(t) \quad \text{c.s.}$$

En los estudios que a continuación realizaremos, trabajaremos, más que con funciones de distribución con funciones de supervivencia, y con lo que llamaremos función de supervivencia aleatoria S, que será la correspondiente a una función de distribución aleatoria F y que vendrá definida como

$$S(t) = 1 - F(t), \qquad \forall t \in [0,\infty) \qquad (2)$$

Por las mismas razones que antes, definiremos a las funciones de supervivencia aleatorias sin tener en cuenta a las correspondientes funciones de distribución aleatorias, aunque las relaciones (1) y (2) deberán ser tenidas siempre presentes para un mejor entendimiento y comprensión de los problemas que estemos considerando.

Definición 1.3.3. Llamaremos función de supervivencia aleatoria a un proceso estocástico $\{S(t) : t \in [0,\infty)\}$, definido sobre algún espacio probabilístico $(\Omega,\sigma_\Omega,\lambda)$ y con espacio de estados el intervalo $[0,1]$, tal que

(1) $S(t)$ es monótona no creciente c.s.

(2) $\lim_{t\to 0} S(t) = 1$ c.s.

(3) $\lim_{t\to +\infty} S(t) = 0$ c.s.

(4) $S(t)$ es contínua por la derecha casi seguramente .

Definición 1.3.4. (Ferguson (1973)): Sea P una medida de probabilidad aleatoria sobre (X,A). Diremos que $X_1,\ldots,X_n$ es una muestra de tamaño n extraida mediante P si para todo $m=1,2,\ldots$ y conjuntos medibles $A_1,\ldots,A_m$, $C_1,\ldots,C_n$ es

$$\Pr\{X_1 \in C_1,\ldots,X_n \in C_n \,/\, P(A_1),\ldots,P(A_m),\ P(C_1),\ldots,P(C_n)\} =$$

$$= \prod_{j=1}^{n} P(C_j) \quad \text{c.s.}$$

en donde Pr denota probabilidad.

De la misma manera, diremos que una muestra de tamaño n es extraida mediante una función de distribución aleatoria, o mediante una función de supervivencia aleatoria, cuando lo es mediante la medida de probabilidad aleatoria asociada.

Teorema 1.3.1. (Kolmogorov (1933)): Cada sistema de funciones de distribución $F_{\mu_1,\mu_2,\ldots,\mu_n}$ satisfaciendo las condiciones (a) y (b) abajo expresadas, define una función de probabilidad $P(A)$ sobre F^M (álgebra de los conjuntos cilíndricos de Borel). Esta función de probabilidad $P(A)$ puede ser extendida, mediante el teorema de extensión a BF^M (mínima σ-álgebra que contiene a F^M).

(a) $F_{\mu_{i_1},\mu_{i_2},\ldots,\mu_{i_n}}(a_{i_1},a_{i_2},\ldots,a_{i_n}) = F_{\mu_1,\mu_2,\ldots,\mu_n}(a_1,a_2,\ldots$
$\ldots a_n)$

(b) $F_{\mu_1\ \mu_2\ \ldots\mu_k}(a_1,a_2,\ldots,a_k) = F_{\mu_1\mu_2\ldots\mu_n}(a_1,a_2,\ldots,a_k,+\ ,\ldots$
$\ldots,+\infty)$

en donde $K < n$ y $\begin{pmatrix} 1, & 2, & \ldots, & n \\ i_1, & i_2, & \ldots, & i_n \end{pmatrix}$ es una permutación arbitraria.

Diremos que $(B_1,\ldots,B_k)$ es una partición medible de X si $B_i \in A\ \forall i$, $B_i \cap B_j = \phi$ para $i \neq j$ y $\bigcup_{j=1}^{k} B_j = X$. Dada una probabilidad aleatoria P, supongamos que hemos definido la distribución conjunta de $(P(B_1),\ldots,P(B_k))$ para cada k y para cada partición medible de X. A partir de esas distribuciones, la distribución conjunta de $(P(A_1),\ldots,P(A_m))$ para conjuntos medibles arbitrarios $A_1,\ldots,A_m$ puede ser definida como sigue, Ferguson (1973):

Dados conjuntos medibles arbitrarios $A_1,\ldots,A_m$, definimos los $k = 2^m$ conjuntos obtenidos tomando intersecciones de cada A_i y sus complementarios; esto es, definimos $B_{\nu_1 \ldots \nu_m}$ para cada $\nu_j = 0$ ó 1 como

$$B_{\nu_1 \ldots \nu_m} = \bigcap_{j=1}^{m} A_j^{\nu_j} \tag{3}$$

en donde A_j^1 es interpretado como A_j y A_j^o como A_j^c, el complementario de A_j. Por tanto, los $\{B_{\nu_1 \ldots \nu_m}\}$ forman una partición de $\mathcal{X}$, y si conocemos ó damos la distribución conjunta de dicha partición, es decir

$$\{P(B_{\nu_1 \ldots \nu_m}); \ \nu_j = 0 \ \text{ó} \ 1, \quad j=1,\ldots,m\} \tag{4}$$

entonces podemos definir la distribución conjunta de $(P(A_1),\ldots$ $\ldots,P(A_m))$ a partir de (4), definiendo para $i=1,\ldots,m$

$$P(A_i) = \sum_{(\nu_1 \ldots \nu_m) \ni \nu_i = 1} P(B_{\nu_1 \ldots \nu_m}) \tag{5}$$

Obsérvese que si $(A_1,\ldots,A_m)$ fuera ya una partición medible al empezar, no se llega a ninguna contradicción con la anterior definición de $(P(A_1),\ldots,P(A_m))$ con tal que $P(\phi)$ sea degenerada en el 0.

Nosotros estamos suponiendo que las distribuciones de las variables aleatorias son definidas libres de su orden, de forma que la condición (a) de Kolmogorov es automática.

Bajo el supuesto de ser $P(\phi)$ degenerada en 0, la distribución de $(P(A_1),\ldots,P(A_m))$ para conjuntos medibles arbitrarios $A_1,\ldots,A_m$ es definida unívocamente, una vez las distribuciones de $(P(B_1),\ldots,P(B_k))$ estén dadas para particiones medibles arbitrarias $(B_1,\ldots,B_k)$.

<u>Condición C. Ferguson (1973)</u>: Si $(B'_1,\ldots,B'_k)$ y $(B_1,\ldots,B_k)$ son particiones medibles, y si $(B'_1,\ldots,B'_k)$ es un refinamiento de

$(B_1,\dots,B_k)$ con $B_1 = \bigcup_{i=1}^{r_1} B'_i$, $B_2 = \bigcup_{i=r_1+1}^{r_2} B'_i,\dots,B_k = \bigcup_{i=r_{k-1}+1}^{k'} B'_i$,

entonces la distribución de

$$\left(\sum_{i=1}^{r_1} P(B'_i),\ \sum_{i=r_1+1}^{r_2} P(B'_i),\dots,\ \sum_{i=r_{k-1}+1}^{k'} P(B'_i)\right)$$

determinada a partir de la distribución conjunta de $(P(B'_1),\dots$ $\dots,P(B'_{k'}))$ es idéntica a la distribución de $(P(B_1),\dots,P(B_k))$.

Teorema 1.3.2. (Ferguson (1973): Si se define un sistema de distribuciones conjuntas de $(P(B_1),\dots,P(B_k))$ para todo k y para toda partición medible $(B_1,\dots,B_k)$ satisfaciendo la condición C, y si para conjuntos medibles arbitrarios $A_1,\dots,A_m$, la distribución de $(P(A_1),\dots,P(A_m))$ se define como en (3), (4) y (5), entonces existe una probabilidad $\mathcal{P}$ sobre $([0,1]^{\mathcal{A}}, BF^{\mathcal{A}})$ produciendo estas distribuciones.

Se ve así que dada una probabilidad aleatoria P, solamente necesitamos dar un sistema de distribuciones para $(P(B_1),\dots,P(B_k))$ para cada partición $(B_1,\dots,B_k)$ y ver si se verifica la condición C, ya que en éste caso se cumplen las condiciones de consistencia de Kolmogorov y por tanto existe la medida de probabilidad (inducida por el proceso) sobre el espacio de las trayectorias $([0,1]^{\mathcal{A}}, BF^{\mathcal{A}})$, en donde $[0,1]^{\mathcal{A}}$ representa el espacio de todas las funciones de $\mathcal{A}$ en $[0,1]$ y $BF^{\mathcal{A}}$ representa la σ-álgebra engendrada por el álgebra de los conjuntos cilíndricos. Dicho conjunto $[0,1]^{\mathcal{A}}$ contiene al conjunto de todas las medidas de probabilidad, el cual es el espacio paramétrico en los problemas de decisión no paramétrica. Vemos así el método para definir medidas de probabilidad a priori en los problemas de decisión Bayesiana no paramétrica: Definimos un proceso. Si ese proceso verifica las condiciones de consistencia de Kolmogorov, entonces existe una medida de probabilidad, que llamaremos a priori, sobre el espacio paramétrico, a la que llamaremos medida de probabilidad inducida por el proceso en cuestión.

Por supuesto, como es

$$F(t) = P((-\infty,t])$$

$$S(t) = 1 - F(t)$$

una vez asegurada la existencia de una medida de probabilidad a priori $\mathcal{P}$ sobre el espacio paramétrico de las medidas de probabilidad P definidas en un espacio muestral dado, tenemos asegurada la existencia de una medida de probabilidad (la misma $\mathcal{P}$) sobre el espacio paramétrico de las funciones de distribución y sobre el espacio paramétrico de las funciones de supervivencia.

Puede ocurrir sin embargo, que el proceso que induzca la medida de probabilidad sea una función de distribución aleatoria o una función de supervivencia aleatoria en vez de una medida de probabilidad aleatoria. En éste caso, en vez de determinar las probabilidades aleatorias correspondientes y ver si cumplen o no la condición C antes establecida, se puede ver si se cumple o no una condición similar de consistencia con respecto a los intervalos de números reales. Más adelante veremos ésto con más detalle al hablar de los procesos neutrales por la derecha.

Vamos ahora con algunos resultados que nos serán de utilidad después.

<u>Definición 1.3.5</u>. Sea $X(t,\omega)$ un proceso estocástico con conjunto de índices T, con espacio de estados R y definido sobre un espacio probabilístico $(\Omega,\sigma_\Omega,\lambda)$. Diremos que $X(t,\omega)$ es contínuo en media cuadrática si y sólo si

$$E[(X(t,\omega) - X(s,\omega))^2] \longrightarrow 0, \quad \text{cuando } s \to t$$

siendo $s,t \in T$, y $\omega \in \Omega$.

A menudo querremos calcular integrales estocásticas en media cuadrática de la forma

$$Y(\omega) = \int_a^b g(t)X(t,\omega)dt$$

pudiendo ser $b = +\infty$, es decir, integrales Lebesgue de la función de t, $g(.)X(.,\omega)$ para ω fijo. Tenemos por tanto una nueva variable aleatoria $Y(\omega)$.

<u>Teorema 1.3.3</u>. (Davis (1977)): Sea $g : [a,b] \longrightarrow R$ una función medible tal que

$$\int_a^b g^2(s)ds < \infty$$

y sea $X(t,\omega)$ un proceso continuo en media cuadrática. Entonces,

$$E\left[(\int_a^b g(t)X(t,\omega)dt).(\int_a^b g(s)X(s,\omega)ds)\right] =$$

$$= \int_a^b \int_a^b g(t).g(s).E[X(t,\omega).X(s,\omega)]dt\ ds.$$

<u>Teorema 1.3.4</u>. Sea $S(t)$, $t \geq 0$ una función de supervivencia aleatoria contínua en media cuadrática. Sea $g : [0,\infty) \longrightarrow R^+ \cup \{0\}$ una función medible no negativa tal que $\forall n \in N$ es

$$\int_0^n g^2(s)ds < \infty.$$

Supongamos que

$$\int_{F'}(\int_0^\infty g(t)S(t)dt)(\int_0^\infty g(u)du)dP(S) < \infty$$

en donde F' es el espacio de todas las funciones de supervivencia y $P(S)$ la probabilidad inducida por el proceso $S(t)$ en dicho espacio. Entonces, es

$$\int_{F'}(\int_0^\infty g(t)S(t)dt)(\int_0^\infty g(u)S(u)du)dP(S) =$$

$$= \int_0^\infty \int_0^\infty g(t)g(u)(\int_{F'} S(t)S(u)dP(S))dtdu$$

En efecto:

$\forall n \in \mathbb{N}$ consideremos las variables aleatorias

$$Y_n = \int_0^n g(t)\ S(t)dt = \int_0^n g(u)S(u)du$$

$$Y = \int_0^\infty g(t)\ S(t)dt = \int_0^\infty g(u)S(u)du$$

Como tanto $g(t)$ como $S(t)$ son funciones no negativas, $\forall n \in \mathbb{N}$ $Y_n \leq Y$ c.s. $\Longrightarrow$ $|Y_n.Y_n| = Y_n.Y_n \leq Y.Y$ c.s.

Por otro lado,

$$\lim_{n\to\infty} Y_n.Y_n = Y.Y \quad \text{c.s.}$$

siendo además $Y.Y$ integrable por hipótesis. Se puede aplicar por tanto el teorema de la convergencia dominada, y será

$$\int Y.Y = \lim_{n\to\infty} \int Y_n\ Y_n$$

Pero por el teorema 1.3.3 es $\forall n \in \mathbb{N}$,

$$\int Y_n \cdot Y_n = \int_0^n \int_0^n g(t)g(u) \left(\int_{F'} S(t)S(u)dP(S)\right)dt\ du$$

con lo que será

$$\int_{F'} \left(\int_0^\infty g(t)S(t)dt\right)\left(\int_0^\infty g(u)S(u)du\right)dP(S) =$$

$$= \int_0^\infty \int_0^\infty g(t)g(u)\left(\int_{F'} S(t)S(u)dP(S)\right)dt\ du$$

y el teorema queda demostrado.

Teorema 1.3.5. Sea $S(t)$, $t \geq 0$ una función de supervivencia aleatoria contínua en media cuadrática. Sean $g_1 : [0,\infty) \longrightarrow R^+ \cup \{0\}$ y $g_2 : [0,\infty) \longrightarrow R^+ \cup \{0\}$ dos funciones medibles no negativas tales que $\forall n \in \mathbb{N}$, es

$$\int_0^n g_1(s)ds < \infty \quad \text{y} \quad \int_0^n g_2(s)ds < \infty.$$

Supongamos que

$$\int_{F'} \left(\int_0^\infty g_i(t)S(t)dt\right)\left(\int_0^\infty g_j(u)S(u)du\right)dP(S) < \infty \qquad \text{para } \begin{array}{l} i=1,2 \\ j=1,2 \end{array}$$

Entonces, es

$$\int_{F'} \left(\int_0^\infty g_1(t)S(t)dt\right)\left(\int_0^\infty g_2(u)S(u)du\right)dP(S) =$$

$$= \int_0^\infty \int_0^\infty g_1(t)g_2(u)\left(\int_{F'} S(t)S(u)dP(S)\right)dt\ du$$

En efecto:

Por el teorema 1.3.4 sabemos que

$$\int_{F'} \left(\int_0^\infty g_1(t)S(t)dt\right)\left(\int_0^\infty g_1(u)S(u)du\right)dP(S) =$$

$$= \int_0^\infty \int_0^\infty g_1(t)g_1(u)\ E\left[S(t)S(u)\right]dt\ du \qquad (6)$$

$$\int_{F'} \left(\int_0^\infty g_2(t)S(t)dt\right)\left(\int_0^\infty g_2(u)S(u)du\right)dP(S) =$$

$$= \int_0^\infty \int_0^\infty g_2(t)g_2(u)\ E\left[S(t)S(u)\right]dt\ du \qquad (7)$$

$$\int_{F'} \left(\int_0^\infty \left[g_1(t)+g_2(t)\right]S(t)dt\right)\left(\int_0^\infty \left[g_1(u)+g_2(u)\right]S(u)du\right)dP(S) =$$

$$= \int_0^\infty \int_0^\infty \left[g_1(t)+g_2(t)\right]\cdot\left[g_1(u)+g_2(u)\right] .\ E\left[S(t).S(u)\right]dt\ du \qquad (8)$$

Pero por otro lado, es

$$\int_{F'} \left(\int_0^\infty \left[g_1(t)+g_2(t)\right]S(t)dt\right)\left(\int_0^\infty \left[g_1(u)+g_2(u)\right]S(u)du\right)dP(S) =$$

$$= \int_{F'} \left(\int_0^\infty g_1(t)S(t)dt\right)\left(\int_0^\infty g_1(u)S(u)du\right)dP(S) +$$

$$+ \int_{F'} \left(\int_0^\infty g_2(t)S(t)dt\right)\left(\int_0^\infty g_2(u)S(u)du\right)dP(S) +$$

$$+ 2\int_{F'} \left(\int_0^\infty g_1(t)S(t)dt\right)\left(\int_0^\infty g_2(u)S(u)du\right)dP(S).$$

Y también,

$$\int_0^\infty \int_0^\infty \left[g_1(t)+g_2(t)\right]\cdot\left[g_1(u)+g_2(u)\right].E\left[S(t).S(u)\right]dt\ du =$$

$$= \int_0^\infty \int_0^\infty g_1(t)g_1(u)\ E\left[S(t).S(u)\right]dt\ du +$$

$$+ \int_0^\infty \int_0^\infty g_2(t)g_2(u)\ E\left[S(t).S(u)\right]dt\ du +$$

$$+ 2\int_0^\infty \int_0^\infty g_1(t)g_2(u)\ E\left[S(t).S(u)\right]dt\ du.$$

Y teniendo en cuenta (6), (7) y (8) deberá de ser

$$2\int_{F'}\left(\int_0^\infty g_1(t)S(t)dt\right)\left(\int_0^\infty g_2(u)S(u)du\right)dP(S) =$$

$$= 2\int_0^\infty\int_0^\infty g_1(t)g_2(u)\ E[S(t).S(u)]dt\ du$$

con lo que el resultado queda probado.

Teorema 1.3.6. Sea $S(t)$, $t \geq 0$ una función de supervivencia aleatoria. Sea $g : [0,\infty) \longrightarrow R$ una función medible de la variable aleatoria T asociada a $S(t)$, tal que

$$\int_{F'}\left(\int_0^\infty |S(t)g(t)|\,dt\right)dP(S) < \infty$$

Entonces,

$$\int_{F'}\left(\int_0^\infty S(t)g(t)dt\right)dP(S) = \int_0^\infty g(t)\left(\int_{F'} S(t)dP(S)\right)dt$$

La demostración es trivial a partir del teorema de Fubini.

Teorema 1.3.7. Sea $S(t)$, $t \geq 0$ una función de supervivencia aleatoria contínua en media cuadrática. Sean $h_1 : [0,\infty) \longrightarrow R$, y $h_2 : [0,\infty) \longrightarrow R$ dos funciones de la variable aleatoria T, tales que sus derivadas con respecto a la medida de Lebesgue sean respectivamente $g_1(x) = \frac{dh_1(x)}{dx}$ y $g_2(y) = \frac{dh_2(y)}{dy}$, las cuales supondremos son funciones medibles no negativas y que verifican la condición

$$\forall n \in N, \qquad \int_0^n g_i(s)ds < \infty, \qquad i=1,2.$$

Supongamos que $c_1 = -\lim_{x\to 0} h_1(x) < \infty$ y que $c_2 = -\lim_{x\to 0} h_2(x) < \infty$.

Supongamos por último que

$$\int_{F'} \left(\int_0^\infty |g_i(t)S(t)|\,dt\right) dP(S) < \infty, \qquad i=1,2$$

y que

$$\int_{F'} \left(\int_0^\infty g_i(t)S(t)\,dt\right)\left(\int_0^\infty g_j(u)S(u)\,du\right) dP(S) < \infty, \quad i=1,2, \quad j=i,2.$$

Entonces,

$$\int_F \left(\int_0^\infty h_1(x)\,dF(x)\right)\left(\int_0^\infty h_2(y)\,dF(y)\right) dP(F) = c_1\cdot c_2 -$$

$$- c_1 \int_0^\infty g_2(y)\left(\int_{F'} S(y)\,dP(S)\right) dy - c_2 \int_0^\infty g_1(x)\left(\int_{F'} S(x)\,dP(S)\right) dx +$$

$$+ \int_0^\infty \int_0^\infty g_1(x) g_2(y)\left(\int_{F'} S(x)S(y)\,dP(S)\right) dx\,dy$$

en donde $F(t) = 1 - S(t)$, $t \geq 0$ es la función de distribución alea toria asociada a $S(t)$.

En efecto:

$$\int_F \left(\int_0^\infty h_1(x)\,dF(x)\right)\left(\int_0^\infty h_2(y)\,dF(y)\right) dP(F) =$$

$$= \int_{F'} \left(\int_0^\infty h_1(x)\,dS(x)\right)\left(\int_0^\infty h_2(y)\,dS(y)\right) dP(S) =$$

$$= \int_{F'} \left(c_1 - \int_0^\infty g_1(x)S(x)\,dx\right)\left(c_2 - \int_0^\infty g_2(y)S(y)\,dy\right) dP(S)$$

integrando por partes, de donde será

$$= c_1\cdot c_2 - c_1 \int_{F'} \left(\int_0^\infty g_2(y)S(y)\,dy\right) dP(S) -$$

$$- c_2 \int_{F'} \left(\int_0^\infty g_1(x)S(x)\,dx\right) dP(S) + \int_{F'} \left(\int_0^\infty g_1(x)S(x)\,dx\right)\left(\int_0^\infty g_2(y)S(y)\,dy\right)\cdot$$

$$\cdot dP(S)$$

de donde aplicando los teoremas 1.3.5 y 1.3.6 obtendremos que es,

$$= c_1 \cdot c_2 - c_1 \int_0^\infty g_2(y)\left(\int_{F'} S(y)dP(S)\right)dy -$$

$$- c_2 \int_0^\infty g_1(x)\left(\int_{F'} S(x)dP(S)\right)dx + \int_0^\infty \int_0^\infty g_1(x)g_2(y)\left(\int_{F'} S(x)S(y)dP(S)\right)$$

$$dx\, dy = c_1 \cdot c_2 - c_1 \int_0^\infty g_2(y)\; E[S(y)]dy -$$

$$- c_2 \int_0^\infty g_1(y)\; E[S(x)]dx + \int_0^\infty \int_0^\infty g_1(x)g_2(y).E[S(x.S(y)]dx\, dy$$

como queríamos demostrar.

<u>Teorema 1.3.8</u>. Sea $S(t)$, $t \geq 0$ una función de supervivencia aleatoria, la cual supondremos induce una medida de probabilidad P sobre F', espacio de todas las funciones de supervivencia. Supongamos que $S(t)$ tiene momentos de órdenes uno y dos con respecto a P. Entonces, la regla Bayes para $S(t)$ con respecto a P, bajo pérdida cuadrática y dentro de la clase de reglas de decisión que son funciones de supervivencia de la forma

$$\hat{S}(t) = a_t\, S_n(t) + b_t\, d(t)$$

en donde $S_n(t)$ es la función de supervivencia muestral definida en 1.2.4, $d(t)$ una función que sólo depende de t y a_t y b_t dos parámetros a determinar, es

$$\hat{S}(t) = \frac{n(E[S^2(t)] - (E[S(t)])^2)}{E[S(t)] + (n-1)E[S^2(t)] - n(E[S(t)])^2}\, S_n(t) +$$

$$+ \frac{E[S(t)] - E[S^2(t)]}{E[S(t)] + (n-1)E[S^2(t)] - n(E[S(t)])^2}\, E[S(t)]$$

en donde todas las esperanzas vienen tomadas con respecto a P; siendo el riesgo Bayes de P,

$$R_{min} = \frac{-(E[S(t)])^3 + E[S^2(t)](E[S(t)])^2 + E[S(t)]E[S^2(t)] - (E[S^2(t)])^2}{E[S(t)] + (n-1)\, E[S^2(t)] - n\,(E[S(t)])^2}$$

Nota: Obsérvese que $\hat{a}_t+\hat{b}_t = 1$, siendo $0 \leq \hat{a}_t \leq 1$ y $0 \leq \hat{b}_t \leq 1$. Obsérvese también que las fórmulas anteriores no son aplicables cuando $E[S^2(t)] = (E[S(t)])^2$, o lo que es lo mismo cuando $V(S(t)) = 0$, pero ésta es la situación en la cual $S(t)$ es una variable aleatoria degenerada, y por tanto aunque caso trivial, carece de importancia como problema no paramétrico, en donde el desconocimiento de la forma de $S(t)$ es característica fundamental. Así pues, en todo éste trabajo supondremos siempre que $V(S(t)) > 0$, aunque como veremos más tarde el valor cero se podrá alcanzar por $V(S(t))$ como situación límite en valores extremos de parámetros que en el intervienen.

En efecto:

Para encontrar la regla Bayes de $S(t)$ deberemos hacer mínimo el riesgo Bayes

$$R = \int_{F'} \int_0^\infty (S(t) - a_t\, S_n(t) - b_t d(t))^2 dQ(S_n(t))dP(S)$$

en donde $Q(S_n(t))$ es la distribución de $S_n(t)$ en el muestreo.

$$R = \int_{F'} S^2(t)dP(S) - 2a_t \int_{F'} S(t) \left(\int_0^\infty S_n(t)dQ(S_n(t))\right)dP(S) -$$

$$- 2b_t\, d(t) \int_{F'} S(t)dP(S) + 2a_t b_t\, d(t)\int_{F'}\left(\int_0^\infty S_n(t)dQ(S_n(t))\right)dP(S)$$

$$+ a_t^2 \int_{F'} \left(\int_0^\infty S_n^2(t)dQ(S_n(t))\right)dP(S) + b_t^2\, d^2(t) =$$

por la proposición 1.2.2,

$$= E_p[S^2(t)]-2a_t\, E_p[S^2(t)] - 2b_t d(t)\, E_p[S(t)] + 2a_t b_t d(t) E_p[S(t)] +$$

$$+ \frac{a_t^2}{n}\,(E_p[S(t)] + (n-1)\, E_p[S^2(t)]) + b_t^2\, d^2(t)$$

con lo que,

$$\left.\begin{aligned} \frac{\partial R}{\partial a_t} &= -2\, E[S^2(t)] + 2b_t d(t)\, E[S(t)] + \frac{2}{n}\, a_t\,(E[S(t)] + (n-1)\, E[S^2(t)]) \\ \frac{\partial R}{\partial b_t} &= -2\, d(t)\, E[S(t)] + 2a_t\, d(t)\, E[S(t)] + 2b_t\, d^2(t) \end{aligned}\right\}$$

en donde la notación se ha simplificado con respecto a las esperanzas. Así pues, igualando dichas derivadas a cero, será

$$a_t = \frac{E[S^2(t)] - b_t\, d(t)\, E[S(t)]}{\frac{1}{n}\,(E[S(t)] + (n-1)\, E[S^2(t)])}$$

y sustituyendolo en la segunda ecuación saldrá,

$$b_t = \frac{(E[S(t)])^2 - E[S(t)]\, E[S^2(t)]}{(E[S(t)] + (n-1)\, E[S^2(t)] - n(E[S(t)])^2)d(t)}$$

con lo que

$$a_t = \frac{n(E[S^2(t)] - (E[S(t)])^2)}{E[S(t)] + (n-1)\, E[S^2(t)] - n(E[S(t)])^2}$$

resultando la regla Bayes por tanto,

$$S(t) = \frac{n(E[S^2(t)] - (E[S(t)])^2)}{E[S(t)] + (n-1)\, E[S^2(t)] - n(E[S(t)])^2}\,.$$

$$\cdot\, S_n(t) + \frac{E[S(t)] - E[S^2(t)]}{E[S(t)] + (n-1)\, E[S^2(t)] - n(E[S(t)])^2}\; E[S(t)]$$

como queríamos demostrar.

En cuanto al riesgo Bayes de P, éste se calculará sustituyendo la regla Bayes obtenida en el riesgo Bayes,

$$R_{min} = E[S^2(t)] - 2\,\frac{n(E[S^2(t)] - (E[S(t)])^2)}{E[S(t)] + (n-1)E[S^2(t)] - n(E[S(t)])^2}\; E[S^2(t)] -$$

$$- 2\,\frac{(E[S(t)])^2 - E[S(t)]\; E[S^2(t)]}{E[S(t)] + (n-1)\; E[S^2(t)] - n(E[S(t)])^2}\; E[S(t)] +$$

$$+ 2n\,\frac{((E[S(t)])^2 - E[S(t)].E[S^2(t)])(E[S^2(t)] - (E[S(t)])^2)}{(E[S(t)] + (n-1)\; E[S^2(t)] - n(E[S(t)])^2)^2}\; E[S(t)] +$$

$$+ n\,\frac{(E[S^2(t)] - (E[S(t)])^2)^2 (E[S(t)] + (n-1)\; E[S^2(t)])}{(E[S(t)] + (n-1)\; E[S^2(t)] - n(E[S(t)])^2)^2} +$$

$$+ \frac{((E[S(t)])^2 - E[S(t)]\; E[S^2(t)])^2}{(E[S(t)] + (n-1)\; E[S^2(t)] - n(E[S(t)])^2)^2} =$$

$$= \frac{n(E[S(t)])^5 - (n\; E[S^2(t)] + 1)(E[S(t)])^4 + 2E[S^2(t)]\; (1-n)(E[S(t)])^3 +}{(E[S(t)] + (n-1)\; E[S^2(t)] - n\; (E[S(t)])^2)^2}$$

$$\frac{+ ((2n-1)(E[S^2(t)])^2 + E[S^2(t)])(E[S(t)])^2 + (n-2)(E[S^2(t)])^2 E[S(t)] +}{}$$

$$\underline{+ (1-n)(E[S^2(t)])^3} =$$

$$= \frac{-(E[S(t)])^3 + E[S^2(t)]\; (E[S(t)])^2 + E[S(t)]\; E[S^2(t)] - (E[S^2(t)])^2}{E[S(t)] + (n-1)\; E[S^2(t)] - n\; (E[S(t)])^2}$$

como queríamos demostrar.

En vista del teorema acabado de demostrar, si realizamos el clá sico procedimiento Bayesiano buscando primero la regla Bayes, bajo pérdida cuadrática, para una función de supervivencia $S(t)$, utilizando una muestra de tamaño n_1, dentro de la clase de reglas de decisión de la forma

$$a_t \; S_{n_1}(t) + b_t \; E[S(t)]$$

es decir, dentro de la clase de reglas de decisión que son combinaciones lineales de la información muestral $(S_{n_1}(t))$ y de la información a priori $(E[S(t)])$, la regla Bayes será

$$\hat{S}_1(t) = \hat{a}_t \; S_{n_1}(t) + \hat{b}_t \; E[S(t)]$$

en donde

$$\hat{a}_t = \frac{n_1(E[S^2(t)] - (E[S(t)])^2)}{E[S(t)] + (n_1-1)E[S^2(t)] - n_1(E[S(t)])^2}$$

y $$\hat{b}_t = \frac{E[S(t)] - E[S^2(t)]}{E[S(t)] + (n_1-1)\; E[S^2(t)] - n_1(E[S(t)])^2}$$

como facilmente se deduce a partir del teorema anterior tomando como $d(t) = E[S(t)]$.

Si a continuación tomamos una muestra de tamaño n_2 y buscamos la regla Bayes para $S(t)$ dentro de la clase de reglas de decisión de la forma

$$a_t \; S_{n_2}(t) + b_t \; \hat{S}_1(t)$$

ésta vendrá dada por la expresión

$$\hat{S}_2(t) = \hat{a}_t \; S_{n_2}(t) + \hat{b}_t \; E[S(t)]$$

sin más que aplicar de nuevo el teorema 1.3.8, tomando $d(t) = \hat{S}_1(t)$, siendo ahora

$$\hat{a}_t = \frac{n_2(E[S^2(t)] - (E[S(t)])^2)}{E[S(t)] + (n_2-1)\ E[S^2(t)] - n_2(E[S(t)])^2}$$

y

$$\hat{b}_t = \frac{E[S(t)] - E[S^2(t)]}{E[S(t)] + (n_2-1)\ E[S^2(t)] - n_2(E[S(t)])^2}$$

Es decir, la única modificación producida ha sido el tamaño muestral, ya que si $n_2 = n_1$ entonces $\hat{S}_1(t) = \hat{S}_2(t)$ no alterando nuestra información a priori $E[S(t)]$ éste nuevo paso, ni disminuyendo el riesgo Bayes salvo en el caso en el que el tamaño muestral $n_2 > n_1$ aunque en dicho caso tampoco dependería el riesgo Bayes del primer paso. Dichos resultados se mantendrían si repitieramos el proceso un número finito de veces. Tenemos por tanto el siguiente resultado.

Teorema 1.3.9. Sea $S(t)$, $t \geq 0$ una función de supervivencia aleatoria como las anteriormente definidas. Si realizamos un procedimiento Bayesiano consistente en extraer una muestra de tamaño n_1 mediante $S(t)$ y en determinar la regla Bayes $\hat{S}_1(t)$, bajo pérdida cuadrática, para $S(t)$ dentro de la clase de reglas de decisión de la forma

$$a_t\ S_{n_1}(t) + b_t\ E[S(t)],$$

a continuación extraer otra muestra de tamaño n_2 y de nuevo determinar la regla Bayes $\hat{S}_2(t)$ dentro de la clase de reglas de decisión de la forma

$$a_t\ S_{n_2}(t) + b_t\ \hat{S}_1(t)$$

y repetir el proceso tomando muestras de tamaños $n_3, n_4, \ldots, n_k$, determinando las reglas Bayes $\hat{S}_3(t)$, $\hat{S}_4(t), \ldots, \hat{S}_k(t)$, buscando respectivamente dentro de la clase de reglas de decisión de la forma

$$a_t \, S_{n_i}(t) + b_t \, \hat{S}_{i-1}(t), \qquad i=3,\ldots,k$$

la regla Bayes resultante es

$$\hat{S}_k(t) = \frac{n_k(E[S^2(t)] - (E[S(t)])^2)}{E[S(t)] + (n_k-1)\, E[S^2(t)] - n_k(E[S(t)])^2} \, S_{n_k}(t) + $$

$$+ \frac{E[S(t)] - E[S^2(t)]}{E[S(t)] + (n_k-1)\, E[S^2(t)] - n_k\,(E[S(t)])^2} \, E[S(t)]$$

siendo el riesgo Bayes resultante,

$$R_{min} = \frac{-(E[S(t)])^3 + E[S^2(t)]\,(E[S(t)])^2 + E[S(t)]\, E[S^2(t)] - (E[S^2(t)])^2}{E[S(t)] + (n_k-1)\, E[S^2(t)] - n_k(E[S(t)])^2}$$

La demostración resulta trivial sin más que aplicar reiteradamente el teorema 1.3.8 k veces, tomando como $d(t) = S_{i-1}(t)$ en la iteración i-ésima, $i=1,\ldots,k$, siendo $S_o(t) = E[S(t)]$ nuestra función de supervivencia a priori, función ésta fundamental en nuestro análisis Bayesiano y sobre la que más tarde volveremos con más detenimiento.

Como se ve, y de ahí la importancia de éste teorema, después de realizar el mencionado proceso Bayesiano, nuestra regla Bayes resultante sólo depende de la última etapa muestral y de nuestra función de supervivencia a priori, y no de las diversas modificaciones que en el transcurso del proceso se hayan ido produciendo, no afectando dicho proceso tampoco al riesgo Bayes correspondiente; de ahí que de ahora en adelante y en vista del teorema, cuando tratemos de estimar la función de supervivencia $S(t)$ solo muestrearemos en una ocasión, ahora eso sí, como el teorema 1.3.10 que más tarde veremos nos aconseja, siendo dicha muestra lo más grande posible.

<u>Teorema 1.3.10</u>. Sea $S(t)$, $t \geq 0$ una función de supervivencia aleatoria como las anteriormente definidas y sea $\hat{S}(t)$ la regla Bayes para $S(t)$ bajo pérdida cuadrática; es decir,

$$\hat{S}(t) = [1 - p_n(t)]\, S_n(t) + p_n(t)\, E[S(t)]$$

en donde

$$p_n(t) = \frac{E[S(t)] - E[S^2(t)]}{E[S(t)] + (n-1)\, E[S^2(t)] - n(E[S(t)])^2} ,$$

y $S_n(t)$ es la función de supervivencia muestral. Entonces $\hat{S}(t)$ converge casi seguro a $S(t)$ cuando $n \to \infty$, para cada $t \geq 0$, y el riesgo Bayes converge hacia cero.

En efecto:

Tenemos que demostrar que

$$P\{|\hat{S}(t) - S(t)| \longrightarrow 0, \quad \text{cuando } n \to \infty\} = 1$$

pero

$$|\hat{S}(t) - S(t)| \overset{c.s.}{=} |(1-p_n(t))S_n(t) + p_n(t)\, E[S(t)] - S(t)| \overset{c.s.}{\leq}$$

$$\overset{c.s.}{\leq} p_n(t)\, |E[S(t)] - S_n(t)| + |S_n(t) - S(t)| \overset{c.s.}{\leq} p_n(t) +$$

$$+ |S_n(t) - S(t)|$$

Por otro lado, como $|S_n(t) - S(t)| \overset{c.s.}{=} |F_n(t) - F(t)|$ por ser $F_n(t) = 1 - S_n(t)$ y $F(t) = 1 - S(t)$ y además ser $|F_n(t) - F(t)| \longrightarrow 0$ cuando $n \to \infty$, c.s., será

$$|S_n(t) - S(t)| \longrightarrow 0, \quad \text{cuando} \quad n \to \infty, \quad \text{c.s.}$$

Además,

$$\lim_{n\to\infty} p_n(t) = 0$$

con lo que

$$|\hat{S}(t) - S(t)| \longrightarrow 0 \quad \text{cuando} \quad n \to \infty, \quad \text{c.s.}$$

y el resultado queda probado.

En cuanto a la convergencia del riesgo, éste puede escribirse como

$$R_{min}(n) = \frac{-(E[S(t)])^3 + E[S^2(t)]\ (E[S(t)])^2 + E[S(t)]\ E[S^2(t)] - (E[S^2(t)])^2}{E[S(t)] - E[S^2(t)] + n(E[S^2(t)] - (E[S(t)])^2)}$$

que claramente tiende a cero cuando $n \to \infty$.

<u>Corolario 1.3.1</u>. La regla Bayes $\hat{S}(t)$ para el parámetro $S(t)$, converge en probabilidad hacia $S(t)$, converge en ley hacia $S(t)$, y la función característica $\phi_{n,t}(\theta)$ asociada al estimador $\hat{S}(t)$, converge hacia la función característica asociada a la desconocida función de supervivencia $S(t)$, cuando $n \to \infty$, y todo esto para cada $t \geq 0$ fijo.

El resultado se obtiene sin más que considerar que la convergencia casi seguro implica la convergencia en probabilidad, ésta la convergencia en ley y ésta última convergencia la de las funciones carac terísticas.

Corolario 1.3.2. En un problema de estimación, el estimador

$$\hat{S}(t) = \left[1 - p_n(t)\right] S_n(t) + p_n(t)\, E\left[S(t)\right]$$

es un estimador consistente para $S(t)$.

Teorema 1.3.11. Sea $S(t)$, $t \geq 0$ una función de supervivencia aleatoria como las anteriormente definidas, la cual supondremos induce una medida de probabilidad P sobre F'. Sea $X = (T_1,\ldots,T_n)$ una muestra aleatoria simple extraida mediante $S(t)$.

Entonces, la regla Bayes para $S(t)$, bajo pérdida cuadrática, con respecto a P, buscada dicha regla dentro del conjunto de todas las reglas de decisión es

$$E_{S(t)/z}\left[S(t)\right]$$

en donde z es un valor de la variable aleatoria $S_n(t)$, la función de supervivencia empírica.

En efecto:

Sea $E_{S(t)/x}\left[S(t)\right]$ la media de la distribución a posteriori, la cual sabemos es una regla Bayes para $S(t)$ bajo pérdida cuadrática. Como $S_n(t)$ es un estimador suficiente para $S(t)$, teorema 1.2.1, podemos aplicar el teorema 1.1.1 ya que toda regla de decisión no aleatorizada (como $E_{S(t)/x}\left[S(t)\right]$) es una regla de decisión aleatorizada, y tendremos que la regla de decisión no aleatorizada

$$d_1 = E_{X/z}\left[E_{S(t)/x}\left[S(t)\right]\right] = E_{S(t)/z}\left[S(t)\right]$$

es R-equivalente a $d_2 = E_{S(t)/x}\left[S(t)\right]$, es decir

$$R(S(t),d_1) = R(S(t),\ d_2)$$

con lo que el riesgo Bayes

$$E_P\left[R(S(t),d_1)\right] = E_P\left[R(S(t),d_2)\right]$$

que será mínimo por ser d_2 regla Bayes para $S(t)$, con lo que el teorema queda probado.

<u>Comparación del estimador lineal $\hat{S}(t)$ y la media a posteriori</u>:

El estimador lineal

$$\hat{S}(t) = \left[1 - p_n(t)\right] S_n(t) + p_n(t)\ E\left[S(t)\right]$$

representa la recta de regresión de $S(t)$ sobre $S_n(t)$ por ser la recta de la forma

$$a\ S_n(t) + b$$

que hace mínimas las desviaciones, teorema 1.3.8.

Por otro, la media a posteriori $E_{S(t)/x}\left[S(t)\right]$ que es la clásica regla Bayes para $S(t)$, tiene el mismo riesgo Bayes que la regla (por tanto también Bayes) $E_{S(t)/z}\left[S(t)\right]$ que representa la curva de regresión de $S(t)$ sobre $S_n(t)$. Así pues, el buscar reglas Bayes lineales como a lo largo de todo éste trabajo haremos no es más que el aproximar la curva regresión por la recta de regresión, y por tanto nuestros estimadores coincidirán con la media a posteriori cuando la recta de regresión y curva lo hagan. En concreto, cuando consideramos como distribución a priori P la inducida por un proceso de Dirichlet así sucede, con lo que podríamos aventurar a los procesos de

Dirichlet en la estadística no paramétrica un papel análogo a la distribución normal en la estadística paramétrica.

Vamos a considerar ahora la situación en la cual hay datos censurados, es decir, algunos de los elementos o individuos de la muestra han muerto, o en general su tiempo de fallo ha sido producido por causas diferentes al estudio que se está realizando; así por ejemplo, si estamos estudiando la muerte por cancer en un grupo de individuos, aquel individuo que muera por otra causa diferente a ésta será considerado como censurado, no debiendo ser eliminado de la muestra porque nos proporciona un dato, su tiempo de fallo (desconocido por haber sido censurado) es posterior al conocido tiempo de censura. Obsérvese que no solamente se consideran censuras las muertes por otras causas, sino también las exclusiones voluntarias de la muestra y cualquier otra causa que evite el determinar con exactitud su tiempo de fallo. Hay que observar también que el número de individuos censurados en una muestra de tamaño n es una variable aleatoria que llamaremos δ y que supondremos independiente de la variable T, siendo $E[\delta] = \delta_o$ conocida. Buena cuestión sería el considerar δ_o desconocida y tratarla como una estimación a priori de δ.

Teorema 1.3.12. Sea $S(t)$, $t \geq 0$ función de supervivencia, desconocida, de la variable aleatoria T. Supongamos que la función de supervivencia aleatoria $S(t)$ induce una medida de probabilidad P sobre F'. Sea $T_1,\ldots,T_n$ una muestra aleatoria simple de T y supongamos que en dicha muestra hay δ individuos censurados, siendo δ independiente de T y $E_P[\delta] = \delta_o$ un número conocido. Supongamos, también que existen los dos primeros momentos de $S(t)$ respecto a P.

Entonces, la regla Bayes para $S(t)$ bajo pérdida cuadrática, con respecto a P, y buscada dicha regla dentro del conjunto de reglas de decisión de la forma

$$a_1[S_n(t)]^{\frac{n-\delta}{n}} + a_2\,[S_n(t)]^{\frac{\delta}{n}} + b\, E[S(t)]$$

es

$$([S_n(t)]^{\frac{n-\delta}{n}}, \quad [S_n(t)]^{\frac{\delta}{n}}, \; E[S(t)]) \quad (\hat{a}_1, \hat{a}_2, \hat{b})'$$

en donde $(\hat{a}_1, \hat{a}_2, \hat{b})'$ viene dado por el sistema

$$K\,(\hat{a}_1, \hat{a}_2, \hat{b})' = \vec{L}$$

viniendo K y $\vec{L}$ especificadas más abajo, y siendo $S_n(t)$ la función de supervivencia muestral.

Así mismo, el riesgo Bayes asociado a P, es

$$R_{min}(n) = E[S^2(t)] - \vec{L}'\,(\hat{a}_1, \hat{a}_2, \hat{b})'$$

En efecto:

Deberemos hacer mínimo el riesgo Bayes

$$R = \int_{F'}\int_0^\infty (S(t) - a_1[S_n(t)]^{\frac{n-\delta}{n}} - a_2[S_n(t)]^{\frac{\delta}{n}} - b\, S_o(t))^2 dQ(S_n(t))dP(S) =$$

$$= \int_{F'} S^2(t)dP(S) + a_1^2 \int_{F'}(\int_0^\infty [S_n(t)]^{2\cdot\left(\frac{n-\delta}{n}\right)} dQ(S_n(t)))dP(S) +$$

$$+ a_2^2 \int_{F'}\int_0^\infty [S_n(t)]^{2\cdot\left(\frac{\delta}{n}\right)} dQ(S_n(t))dP(S) + b^2(E[S(t)])^2 -$$

$$- 2a_1 \int_{F'} S(t)(\int_0^\infty [S_n(t)]^{\frac{n-\delta}{n}} dQ(S_n(t)))dP(S) -$$

$$- 2a_2 \int_{F'} S(t)(\int_0^\infty [S_n(t)]^{\frac{\delta}{n}} dQ(S_n(t)))dP(S) -$$

$$- 2b\ E[S(t)] \int_{F'} S(t)dP(S) + 2a_1a_2 \int_{F'}(\int_0^\infty S_n(t)dQ(S_n(t)))dP(S) +$$

$$+ 2a_1\ b\ E[S(t)] \int_{F'}(\int_0^\infty [S_n(t)]^{\frac{n-\delta}{n}} dQ(S_n(t)))dP(S) +$$

$$+ 2a_2\ b\ E[S(t)] \int_{F'}(\int_0^\infty [S_n(t)]^{\frac{\delta}{n}} dQ(S_n(t)))dP(S)$$

Por otro lado, $[S_n(t)]^a = [1 - F_n(t)]^a$, y el desarrollo en series de potencias de $(1-x)^a$ es

$$(1-x)^a = 1-ax + a(a-1)\frac{x^2}{2} - \ldots$$

con lo que si cogemos los dos primeros términos del desarrollo será

$$[S_n(t)]^{\frac{\delta}{n}} = 1 - \frac{\delta}{n} F_n(t) = 1 - \frac{\delta}{n} + \frac{\delta}{n} S_n(t)$$

$$[S_n(t)]^{1-\frac{\delta}{n}} = 1 - (1 - \frac{\delta}{n})F_n(t) = \frac{\delta}{n} + (1 - \frac{\delta}{n})\ S_n(t)$$

$$[S_n(t)]^{2\frac{\delta}{n}} = 1 - 2\frac{\delta}{n} F_n(t) = 1 - \frac{2\delta}{n} + \frac{2\delta}{n} S_n(t)$$

$$[S_n(t)]^{2(1-\frac{\delta}{n})} = 1 - 2(1 - \frac{\delta}{n})\ F_n(t) = -1 + \frac{2\delta}{n} + (2 - \frac{2\delta}{n})\ S_n(t)$$

con lo que como

$E[S_n(t)] = S(t)$, $\quad E[S_n^2(t)] = \frac{S(t)}{n} + \frac{n-1}{n} S^2(t)$ $\quad$ y

$E[\delta] = \delta_o$ $\quad$ será,

$$R = E[S^2(t)] + a_1^2(-1 + \frac{2}{n}\delta_o + (1 - \frac{2\delta_o}{n})E[S(t)]) +$$

$$+ a_2^2(1 - \frac{2}{n}\delta_o + \frac{2}{n}\delta_o\ E[S(t)]) + b^2(E[S(t)])^2 -$$

$$- 2a_1(\frac{\delta_o}{n} E[S(t)] + (1 - \frac{\delta_o}{n}) E[S^2(t)]) -$$

$$- 2a_2((1 - \frac{\delta_o}{n}) E[S(t)] + \frac{\delta_o}{n} E[S^2(t)]) - 2b(E[S(t)])^2 +$$

$$+ 2a_1 E[S(t)] + 2a_1 b (\frac{\delta_o}{n} E[S(t)] + (1 - \frac{\delta_o}{n})(E[S(t)])^2) +$$

$$+ 2a_2 b ((1 - \frac{\delta_o}{n}) E[S(t)] + \frac{\delta_o}{n} (E[S(t)])^2)$$

y derivando, se obtiene el llamado sistema de ecuaciones normales,

$$K \vec{\beta} = \vec{L}$$

en donde la matriz K es la dada por

$$K = \begin{bmatrix} -1 + \frac{2}{n}\delta_o + 2\left(1 - \frac{\delta_o}{n}\right)E[S(t)] & E[S(t)] & \frac{\delta_o}{n} E[S(t)] + \left(1 - \frac{\delta_o}{n}\right)\left(E[S(t)]\right)^2 \\ E[S(t)] & 1 - \frac{2}{n}\delta_o + \frac{2}{n}\delta_o E[S(t)] & \left(1 - \frac{\delta_o}{n}\right)E[S(t)] + \frac{\delta_o}{n} E^2[S(t)] \\ \frac{\delta_o}{n} E[S(t)] + \left(1 - \frac{\delta_o}{n}\right)\left(E[S(t)]\right)^2 & \left(1 - \frac{\delta_o}{n}\right)E[S(t)] + \frac{\delta_o}{n}\left(E[S(t)]\right)^2 & \left(E[S(t)]\right)^2 \end{bmatrix}$$

$\vec{L}$ el dado por

$$\vec{L} = (\frac{\delta_o}{n} E[S(t)] + (1 - \frac{\delta_o}{n}) E[S^2(t)], \ (1 - \frac{\delta_o}{n}) E[S(t)] + \frac{\delta_o}{n} E[S^2(t)],$$
$$(E[S(t)])^2)'$$

y $\vec{\beta} = (a_1, a_2, b)'$ el vector de parámetros,

con lo que el riesgo mínimo quedará

$$R_{min}(n) = E[S^2(t)] - 2\,\hat{\vec{\beta}}'\,\vec{L} + \hat{\vec{\beta}}'\,K\,\hat{\vec{\beta}}$$

como queríamos demostrar.

Vamos a definir a continuación una nueva función aleatoria que más tarde utilizaremos.

Definición 1.3.6. Sea $\beta(t)$, $t \geq 0$ una función contínua real valorada positiva acotada siempre por cero y con límites a la izquierda existentes.

Sea $Z(t)$, $t \geq 0$ un proceso con incrementos independientes definido sobre un apropiado espacio probabilístico (Ω, A, P); es decir, $Z(0) = 0$ y $Z(t)$ tiene incrementos independientes. Supondremos que éste proceso tiene trayectorias contínuas por la derecha no decrecientes.

Llamaremos función tasa de azar aleatoria, correspondiente a la función $\beta(t)$ y al proceso $Z(t)$, al proceso estocástico

$$r(t) = \int_{[0,t)} \beta(s)\,dZ(s)$$

en donde la integración se realiza con respecto a las trayectorias del proceso $Z(t)$.

Definición 1.3.7. Sea $r(t)$ una función tasa de azar aleatoria, correspondiente a la función $\beta(t)$ y al proceso $Z(t)$. Llamaremos función de azar acumulativa aleatoria correspondiente a $r(t)$ a la función

$$H(t) = \int_{[0,t)} r(s)ds$$

<u>Teorema 1.3.13</u>. Sea $H(t)$ una función de azar acumulativa aleatoria con función $\beta(t)$ y proceso $Z(t)$ asociados. Entonces, el proceso

$$S(t) = e^{-H(t)}, \quad t \geq 0$$

es una función de supervivencia aleatoria.

El resultado se sigue de la comprobación inmediata de las propiedades que caracterizan a una función de supervivencia aleatoria, a partir de las propiedades de una tasa de azar acumulativa aleatoria.

IV. <u>PROCESOS DE DIRICHLET</u>

Dentro de la decisión Bayesiana no paramétrica, la primera medida de probabilidad a priori P considerada sobre el espacio paramétrico, el cual en los problemas no paramétricos es el espacio de todas las distribuciones de probabilidad definidas sobre un espacio muestral dado, fue la inducida por un proceso de Dirichlet P. Dichos procesos, los cuales son una clase de probabilidades aleatorias, fueron introducidos por Ferguson en 1973.

Estos procesos son hoy por hoy, las probabilidades a priori más sencillas dentro de los problemas no paramétricos.

Las principales propiedades o características de dichos procesos son que,

(1) P es no paramétrica en el sentido de que tiene una clase de probabilidades "grande" o "no paramétrica" como sopor

te suyo en la topología de la convergencia débil.

(2) Si P es considerado como un parámetro con distribución a priori $\mathcal{P}$, entonces la distribución a posteriori de P, dada una muestra, también tiene una distribución de Dirichlet.

(3) P es una probabilidad discreta con probabilidad uno.

<u>La distribución gamma</u>:

Utilizaremos la notación $X \equiv G(a,b)$ para indicar que la variable aleatoria X tiene una distribución gamma con parámetro de forma $a \geq 0$ y parámetro de escala $b > 0$. Los convenios que adoptaremos son que $G(0,b)$ representa a la distribución degenerada en el punto 0, es decir, a aquella con masa unidad en el punto 0, y que $G(\infty-\infty,b)$ o $G(\infty,b)$ representan a la distribución con masa uno en ∞.

Si $a > 0$ ésta distribución tiene densidad con respecto a la medida de Lebesgue en la recta real, siendo ésta

$$f(x/a,b) = \frac{b^a \, e^{-bx} \, x^{a-1}}{\Gamma(a)} \cdot I_{(o,\infty)}(x)$$

en donde $I_S(z)$ representa a la función indicador del conjunto S.

Si $X \equiv G(a,b)$, entonces es

$$E[X] = \frac{a}{b} \qquad \text{y} \qquad V(X) = \frac{a}{b^2}$$

La distribución de Dirichlet:

Sean $X_1,\ldots,X_n$ n variables aleatorias independientes, de forma que cada variable X_j sigue una distribución gamma $G(\alpha_j,1)$, para $j=1,\ldots,n$ y con $\alpha_j \geq 0$, existiendo algún i, de forma que $\alpha_i > 0$.

Consideremos las variables aleatorias

$$Y_j = \frac{X_j}{\sum_{k=1}^{n} X_k}, \qquad j=1,2,\ldots,n$$

A la distribución de la variable aleatoria n-dimensional $(Y_1,\ldots,Y_n)$ se le denomina distribución de Dirichlet de parámetros $(\alpha_1,\ldots,\alpha_n)$, la cual representaremos por $\mathcal{D}(\alpha_1,\ldots,\alpha_n)$.

Si $\alpha_j > 0$ para $j=1,\ldots,n$, la variable aleatoria (n-1)-dimensional $(Y_1,\ldots,Y_{n-1})$ es absolutamente contínua con respecto a la medida de Lebesgue en R^{n-1} y tiene por función de densidad

$$f(y_1,\ldots,y_{n-1}/\alpha_1,\ldots,\alpha_n) = \frac{\Gamma(\alpha_1+\ldots+\alpha_n)}{\Gamma(\alpha_1)\ \ldots\ \Gamma(\alpha_n)} \left(\prod_{j=1}^{n-1} y_j^{\alpha_j-1}\right) .$$

$$. \left(1-\sum_{j=1}^{n-1} y_j\right)^{\alpha_n-1} . \ I_S(y_1,\ldots,y_{n-1})$$

en donde S es el simplex,

$$S = \{(y_1,\ldots,y_{n-1}) : y_i \geq 0, \ i=1,\ldots,n-1, \ \sum_{i=1}^{n-1} y_i \leq 1\}$$

Proposición 1.4.1. Si $(Y_1,\ldots,Y_k) \in \mathcal{D}(\alpha_1,\ldots,\alpha_k)$ y si $r_1,\ldots,r_e$ son enteros tales que $0 < r_1 < \ldots < r_e = k$, entonces,

$$\left(\sum_{i=1}^{r_1} Y_i, \sum_{i=r_1+1}^{r_2} Y_i, \ldots, \sum_{i=r_{e-1}+1}^{r_e} Y_i\right) \in$$

$$\in \mathcal{D}\left(\sum_{i=1}^{r_1} \alpha_i, \sum_{i=r_1+1}^{r_2} \alpha_i, \ldots, \sum_{i=r_{e-1}+1}^{r_e} \alpha_i\right)$$

La demostración se sigue inmediatamente de la definición de la distribución de Dirichlet y de la reproductividad respecto al primer parámetro de la distribución gamma.

Para más propiedades de ésta distribución puede consultarse Wilks (1962).

Procesos de Dirichlet. Definición. Ferguson (1973):

Sea (X,A) un espacio medible y sea α una medida finita no nula sobre (X,A). Definimos un proceso estocástico particular $\{P(A) : A \in A\}$ de la siguiente manera: P es un proceso de Dirichlet sobre (X,A) con parámetro α, si para cada $k=1,2,\ldots$ y para cada partición medible $(B_1,\ldots,B_k)$ de X, la distribución de la variable aleatoria k-dimensional $(P(B_1),\ldots,P(B_k))$ es de Dirichlet con parámetros $(\alpha(B_1,\ldots,\alpha(B_k))$.

Claramente, la condición C de consistencia de Ferguson se sigue inmediatamente de la proposición 1.4.1, con lo que por el teorema 1.3.1, sabemos que existe una medida de probabilidad a priori sobre el espacio paramétrico, considerando en el peor de los casos una versión separable de P.

<u>Muestra extraida mediante un proceso de Dirichlet. Ferguson (1973)</u>:

La colección de variables aleatorias $X_1,\ldots,X_n$ con valores sobre $(\mathcal{X},\mathcal{A})$ se dice que es una muestra de tamaño n obtenida a través de un proceso de Dirichlet P sobre $(\mathcal{X},\mathcal{A})$ con parámetro α, si para todo $m=1,2,\ldots$ y conjuntos $\mathcal{A}$-medibles $A_1,\ldots,A_m,C_1,\ldots,C_n$ es

$$Pr\{X_1 \in C_1,\ldots,X_n \in C_n / P(A_1),\ldots,P(A_m),P(C_1),\ldots,P(C_n)\} =$$

$$= \prod_{j=1}^{n} P(C_j) \qquad \text{c.s.}$$

en donde P_r denota probabilidad.

Como se ve, esta definición no es más que un caso particular de la definición 1.3.4., por ser un proceso de Dirichlet un caso particular de probabilidad aleatoria.

<u>Teorema 1.4.1</u>. Ferguson (1973): Sea P un proceso de Dirichlet sobre $(\mathcal{X},\mathcal{A})$ con parámetro α, y sea $X_1,\ldots,X_n$ una muestra de tamaño n extraida mediante P. Entonces, la distribución condicional de P dada $X_1,\ldots,X_n$ es un proceso de Dirichlet de parámetro $\alpha + \sum_{i=1}^{n} \delta_{X_i}$, en donde para cada $x \in \mathcal{X}$, δ_x denota la medida sobre $(\mathcal{X},\mathcal{A})$ que da masa uno al punto x:

$$\delta_x(A) = \begin{cases} 1 & \text{si} \quad x \in A \\ 0 & \text{si} \quad x \notin A \end{cases}$$

Como se ve, la distribución a posteriori de un proceso de Dirichlet es un proceso de Dirichlet muy facilmente manejable, lo que permite estimar con facilidad, ya que por ejemplo, si consideramos la pérdida cuadrática $L(\theta,a) = (\theta-a)^2$, las reglas Bayes van a ser las medias con respecto a esa distribución a posteriori tan manejable.

V. PROCESOS NEUTRALES POR LA DERECHA

Otras probabilidades a priori que se pueden considerar sobre el espacio paramétrico de todas las distribuciones de probabilidad definidas en un espacio muestral dado, son los procesos neutrales por la derecha, introducidos por Doksum en 1974.

Estos procesos gozan de propiedades semejantes a los procesos de Dirichlet en el sentido de ser "no paramétricos" y de ser la distribución a posteriori de una probabilidad aleatoria, neutral a la derecha. De hecho, como luego veremos, los procesos de Dirichlet son un caso particular de proceso neutral por la derecha.

Definición 1.5.1. Diremos que la función de distribución aleatoria $F(t)$ es un proceso neutral por la derecha, si puede ser escrito de la forma

$$F(t) = 1 - e^{-Y_t}$$

en donde Y_t es un proceso con incrementos independientes, tal que

1) Y_t es no decreciente c.s.

2) Y_t es contínuo por la derecha c.s.

3) $\lim_{t\to -\infty} Y_t = 0$ c.s.

4) $\lim_{t\to +\infty} Y_t = \infty$ c.s.

Intuitivamente, una distribución aleatoria $F(t)$ es un proceso neutral por la derecha si para cada $t_1, t_2 \in R$ con $t_1 < t_2$,

$$\frac{1-F(t_2)}{1-F(t_1)} \quad \text{es independiente de} \quad \{F(t) : t \leq t_1\}$$

es decir, si la porción de masa que $F(t)$ asigna al subintervalo (t_2,∞) del intervalo (t_1,∞) es independiente de lo que valga $F(t)$ a la izquierda de t_1.

Claramente, diremos que la función de supervivencia aleatoria $S(t)$ es un proceso neutral por la derecha si $F(t) = 1 - S(t)$ es una función de distribución aleatoria la cual es un proceso neutral por la derecha.

Si consideramos la situación, que será en la que estaremos a lo largo de todo éste trabajo, de ser $t \geq 0$ y consideramos una partición de $[0,\infty)$ en un número finito k de intervalos disjuntos $[a_o = 0, a_1]$, $(a_1,a_2], \ldots, (a_{k-1}, a_k = \infty)$ y llamamos $r_i = Y_{a_i} - Y_{a_{i-1}}$, Doksum (1974), a partir de los resultados de Ferguson (1973), ha visto que se puede especificar una distribución de probabilidad sobre el espacio de las trayectorias del proceso Y_t (y por tanto sobre el espacio paramétrico en los problemas de decisión Bayesiana no paramétrica), especificando las distribuciones finito dimensionales de las r_i para cada partición $(a_{i-1}, a_i]$, $i=1,\ldots,k$, sujetas a la condición de que la distribución de $r_i + r_{i+1}$ debe de ser la misma que la obtenida por aplicación directa de las reglas al intervalo combinado $(a_{i-1}, a_{i+1}]$. Pero dado que Y_t es un proceso con incrementos independientes, ésto siempre se va a verificar, pues las r_i son los incrementos del proceso.

Así pues, el problema de la existencia de una distribución de probabilidad a priori P, que llamaremos inducida por el proceso, sobre el espacio paramétrico en los problemas de decisión Bayesiana no paramétrica, se reduce a la especificación del proceso con incrementos independientes Y_t, o alternativamente al de especificar las distribuciones de las variables aleatorias r_i para cada partición finita $(a_{i-1},a_i]$, $i=1,\ldots,k$, sujetas a la condición de consistencia arriba especificada.

Hay que observar que dado un proceso neutral por la derecha $F(t)$, puede suceder que el espacio paramétrico no sea el espacio de funciones de distribución no aleatorias, pero como Doksum (1974) dice, existe una versión separable de él que si lo verifica, es decir existe una versión separable de $F(t)$ tal que $P\{F : F$ es una función de distribución$\} = 1$, en donde P es la inducida por el proceso $F(t)$. También es bueno observar que P no está definida solo sobre el conjunto de funciones de distribución, pero podemos decir que P elige una $F(t)$ que se puede "arreglar" mediante su versión separable, que sabemos siempre existe por un teorema debido a Doob (1953), de forma que sea ésta una función de distribución; así pues, supondremos se verifican todas las condiciones necesarias para considerar a ésta P como una medida de probabilidad a priori en el sentido tradicional y clásico, que en cualquier caso siempre podremos suponer existente. Por último hacer notar que cuando calculemos esperanzas respecto a ésta P estaremos integrando en un conjunto más amplio que F, pero por razones plásticas de similitud con el caso paramétrico, utilizaremos una notación de la forma $\int_F \ldots dP(F)$.

"

Muestra extraida mediante un proceso neutral por la derecha.

DOCKSUM (1974):

Sea $\mathbb{R}$ la recta real y β la σ-álgebra de Borel. Sea $F(t)$ un proceso neutral por la derecha con distribución inducida P. Sea $(X_1,\ldots,X_n)$ un vector aleatorio en $\mathbb{R}^n$. Diremos que la distribución condicionada de $(X_1,\ldots,X_n)$ dado F es la de una muestra aleatoria obtenida a través de una distribución F. Esta distribución condicionada puede encontrarse del modo usual definiendo la distribución conjunta P_2 del vector $(X_1,\ldots,X_n)$ y de F como sigue:

$$P_2(F \in D,\ X_1 \in (-\infty,t_1],\ldots,X_n \in (-\infty,t_n]) =$$

$$= E_P\left[I_D(F)\prod_{i=1}^{n} F(t_i)\right]$$

en donde D está en $\sigma(\beta^\beta)$, $(-\infty,t_i] \in \beta$ e I_D es la función indicatriz de D.

Entonces, P_2 es una probabilidad sobre $(\mathbb{R}^n \times [0,1]^\beta,\ \sigma(\beta^n \times \beta^\beta))$. Con esta definición observamos que

1) La distribución marginal de F es P.

2) $P_2(X_1 \in (-\infty,t_1],\ldots,X_n \in (-\infty,t_n]/F) = \prod_{i=1}^{n} F(t_i)$ c.s.

3) La distribución marginal de $(X_1,\ldots,X_n)$ es

$$P_2(X_1 \in (-\infty,t_1],\ldots,X_n \in (-\infty,t_n]) = E_P\left[\prod_{i=1}^{n} F(t_i)\right].$$

en particular

$$P_2(X \in (-\infty,t]) = E_P[F(t)] = \nu(t)$$

Doksum (1974) demuestra que $\nu(t)$ es una distribución y que ca racteriza al proceso neutral por la derecha. A dicha distribución $\nu(t)$ se le llama parámetro del proceso.

Entre las propiedades más notables de los procesos neutrales por la derecha podemos citar la propiedad Bayesiana establecida por el siguiente teorema.

Teorema 1.5.1. (Doksum (1974): Si F es un proceso neutral por la derecha, entonces la distribución a posteriori de F dado $X_1,\ldots,X_n$ es la inducida por un proceso neutral por la derecha.

Sin embargo, a diferencia de los procesos de Dirichlet, la distribución a posteriori de un proceso neutral por la derecha resulta inmanejable en general, a la hora de hacer estimaciones. Es por ésto por lo que posteriormente propondremos un método para poder hacer estimaciones con algunos procesos a la derecha particulares.

Distribución marginal de la muestra:

Dada una muestra X extraida mediante un proceso neutral por la derecha $F(t)$, la distribución marginal de la muestra será

$$\nu(t) = \int_F F(t)\, dP(F)$$

que es el parámetro del proceso nuetrla por la derecha, con lo que

$$d\nu(t) = \int_F dF(t)dP(F) \qquad (9)$$

Medida de Lévy asociada a un proceso neutral por la derecha:

El proceso Y_t descrito en la definición 1.5.1, tiene a lo sumo una cantidad numerable de puntos fijos de discontinuidad que podemos llamar $t_1, t_2, \ldots,$ en algún orden. Sean $S_1, S_2, \ldots,$ los tamaños aleatorios de los saltos de Y_t en $t_1, t_2, \ldots,$ respectivamente. Así pues, $S_1, S_2, \ldots,$ son variables aleatorias independientes no negativas.

La diferencia

$$Z_t = Y_t - \sum_{j=1}^{\infty} S_j \, I_{[t_j,\infty)}(t)$$

en donde I_B representa a la función indicador del conjunto B, es un proceso con incrementos independientes, no decreciente c.s., sin puntos fijos de discontinuidad, siendo $\lim_{t \to -\infty} Z_t = 0$ c.s.. Debe de tener por tanto Z_t una distribución infinitamente divisible.

Para un proceso con incrementos independientes y sin puntos fijos de discontinuidad, la representación de Lévy-Kolmogorov para el logaritmo de la función generatriz de momentos (obsérvese el signo menos), es de la forma

$$\log E\left[e^{-\theta Z_t}\right] = -\theta\, \alpha(t) + \int_0^{\infty} (e^{-\theta u} - 1 + \theta u) \, \frac{dK_t(u)}{u^2} \tag{10}$$

siempre que $\mathrm{Var}(Z_t) < \infty$; siendo $\alpha(t) = E[Z_t]$ y K_t una función sobre $\mathbb{R}$ contínua a la derecha, no decreciente y acotada tal que $K_t(-\infty) = 0$ y $K_t(+\infty) < \infty$.

Observese como los procesos neutrales a la derecha no tienen parte aleatoria por ser Z_t un proceso con incrementos independientes, Ferguson y Klass (1972).

A partir de (10) podemos escribir,

$$\log E\left[e^{-\theta Z_t}\right] = -\theta\left[\alpha(t) - \int_0^\infty \frac{dK_t(u)}{u}\right] + \int_0^\infty (e^{-\theta u} - 1)\, \frac{dK_t(u)}{u^2}$$

o lo que es lo mismo,

$$\log E\left[e^{-\theta Z_t}\right] = -\theta b(t) + \int_0^\infty (e^{-\theta u} - 1)\ d\,N_t(u)$$

siendo $b(t) = \alpha(t) - \int_0^\infty \frac{dK_t(u)}{u}$ y $d\,N_t(u) = \frac{dK_t(u)}{u^2}$.

A la medida $N_t(u)$ la llamaremos medida de Lévy asociada al proceso neutral a la derecha $F(t)$. Ahora la parte no aleatoria viene representada por $b(t)$.

Para que exista el momento de segundo orden de Z_t y equivalentemente $V(Z_t)$, es necesario y suficiente, Feller (1966), que la medida $K_t(u)$ sea finita, es decir que

$$\int_0^\infty dK_t(u) < \infty$$

y por ser $d\,N_t(u) = \frac{dK_t(u)}{u^2}$, necesario y suficiente que

$$\int_0^\infty u^2\ dN_t(u) < \infty$$

Por otro lado, la medida $N_t(u)$ verifica

$$\int_0^\infty \frac{u}{1+u}\ dN_t(u) < \infty$$

por estar Z_t concentrada en $(0,\infty)$, Feller (1966).

Así pues, asociada al proceso neutral por la derecha $F(t)$ existe una medida, $\forall t \in R$ $N_t(u)$, llamada de Lévy, sobre los subconjuntos de Borel de $(0,\infty)$ que verifica

$$\int_0^\infty \frac{z}{1+z}\ dN_t(t) < \infty$$

y que para que (10) tenga sentido, es decir para que exista $V(Z_t)$ debe de cumplir

$$\int_0^\infty z^2\ dN_t(z) < \infty$$

Por último, obsérvese que cuando Y_t no tenga puntos fijos de discontinuidad, Y_t coincidirá con Z_t.

<u>El proceso de Dirichlet como caso particular de P.N.D.:</u>

Al proceso de Dirichlet le corresponde un proceso con incrementos independientes Y_t que no tiene puntos fijos de discontinuidad (por lo que coincidirán Y_t y Z_t), ni parte no aleatoria, es decir, $b \equiv 0$, siendo el logaritmo de la función generatriz de momentos por tanto,

$$\log E\left[e^{-\theta Y_t}\right] = \int_0^\infty (e^{-\theta z} - 1)\ dN_t(z)$$

en donde la medida de Lévy asociada es la dada por

$$dN_t(z) = \frac{e^{-\alpha(R)z}\ (e^{\alpha(t)z} - 1)}{Z\ .\ (1-e^{-z})}\ dt$$

y en donde α es el parámetro del proceso de Dirichlet. Además, si P es un proceso de Dirichlet de parámetro α, la distribución $\nu(t)$, parámetro del proceso neutral por la derecha, es

$$\nu(t) = \frac{\alpha(t)}{\alpha(R)}$$

El proceso homogéneo neutral a la derecha:

Un tipo muy especial de proceso neutral a la derecha es el proceso homogéneo. Diremos que una función de distribución aleatoria $F(t)$ neutral a la derecha es un proceso homogéneo si el proceso con incrementos independientes correspondiente $Y_t = -\log(1 - F(t))$ no tiene puntos fijos de discontinuidad ni parte no aleatoria, y su medida de Lévy asociada es independiente de t, es decir de la forma $N_t(.) = \gamma(t)N(.)$. Así pues, el logaritmo de la función generatriz de momentos es de la forma

$$\log E\left[e^{-\theta Y_t}\right] = \gamma(t) \int_0^\infty (e^{-\theta z} - 1)\, dN(z)$$

en donde $\gamma(t)$ es una función contínua no decreciente con $\lim_{t\to -\infty} \gamma(t) = 0$, $\lim_{t\to +\infty} \gamma(t) = +\infty$, y en donde N es una medida sobre $(0,\infty)$ tal que

$$\int_0^\infty \frac{z}{1+z}\, dN(t) < \infty.$$

Próximamente veremos con más detalle dos de éstos procesos homogéneos, el proceso gamma exponencial y el proceso homogéneo simple.

Por último hacer notar que, como claramente se ve, el proceso de Dirichlet no es un proceso homogéneo, es decir, es un proceso no homogéneo.

Proposición 1.5.1. (Doksum (1974)): Si una distribución aleatoria $F(t) = 1 - e^{-Y_t}$ es un proceso neutral a la derecha y si la correspondiente $Y_t = -\log\left[1 - F(t)\right]$ no tiene parte no aleatoria, entonces

$\mathcal{P}\{F : F$ es una función de distribución discreta$\} = 1$,

siendo $\mathcal{P}$ la probabilidad inducida por el proceso.

Así pues, en un problema de decisión Bayesiana no paramétrico, si consideramos como distribución a priori un proceso neutral a la derecha sin parte no aleatoria, extraeríamos de una supuesta urna que contuviese todas las distribuciones de probabilidad, una discreta con probabilidad uno. En particular este resultado se mantiene para los procesos de Dirichlet y para los procesos homogéneos, como los gamma exponenciales y los homogéneos simples.

VI. PROCESOS GAMMA EXPONENCIALES

Sea $T \geq 0$ una variable aleatoria observable la cual nos indica el tiempo de fallo de un sistema.

Consideremos la función de supervivencia aleatoria

$$S(t/\Lambda_o) = e^{-\Lambda_o(t)}, \quad t \geq 0 \tag{11}$$

que nos indica que

$$\Pr\{T > t/\Lambda_o\} = S(t/\Lambda_o)$$

en donde $\Lambda_o(t)$ es un proceso estocástico que especificaremos a través de la definición de la función de supervivencia aleatoria. Obsérvese que esta probabilidad es la probabilidad aleatoria asociada.

Se cumple que,

(1) $\lim_{t \to o} \Lambda_o(t) = 0$ c.s.,

(2) $\lim\limits_{t \to \infty} \Lambda_o(t) = \infty$ c.s.,

(3) $\Lambda_o(t) - \Lambda_o(s) \geq 0$ c.s., para $t > s$

para que se cumplan respectivamente las condiciones (2), (3) y (1) de la definición 1.3.3.

Consideremos una partición de $[0,\infty)$ en un número finito de intervalos disjuntos:

$$[a_o = 0, a_1], \quad (a_1, a_2], \ldots, (a_{k-1}, a_k = \infty)$$

Observemos que el primer intervalo es cerrado y el último es abierto, aunque cuando tratemos de resumirlos todos, por la notación pueda parecer abierto también el primero. Si denotamos por,

$$q_i = \begin{cases} \Pr\{T \in (a_{i-1}, a_i] / T > a_{i-1}, \Lambda_o\} & \text{si} \quad \Pr\{T > a_{i-1}/\Lambda_o\} > 0 \\ \\ 1 & \text{en otro caso} \end{cases}$$

$i=1,2,\ldots,k$

tenemos la siguiente proposición.

<u>Proposición 1.6.1.</u>

$$\Lambda_o(a_i) = \sum_{j=1}^{i} -\log(1 - q_j) = \sum_{j=1}^{i} r_j, \quad i=1,\ldots,k.$$

Es decir, en cada punto a_i, $i=1,\ldots,k$, $\Lambda_o(a_i)$ es la suma de i variables aleatorias no negativas.

En efecto: Si $\Pr\{T > a_{i-1}/\Lambda_o\} > 0$, para $i=1,2,\ldots,k$,

$$q_i = \Pr\{T \in (a_{i-1}, a_i] / T > a_{i-1}, \Lambda_o\} =$$

$$= \frac{\Pr\{T \in (a_{i-1}, a_i] \cap T > a_{i-1} / \Lambda_o\}}{\Pr\{T > a_{i-1} / \Lambda_o\}} = \frac{\Pr\{T \in (a_{i-1}, a_i] / \Lambda_o\}}{\Pr\{T > a_{i-1} / \Lambda_o\}}$$

$$= \frac{S(a_{i-1}/\Lambda_o) - S(a_i/\Lambda_o)}{S(a_{i-1}/\Lambda_o)}$$

y sustituyendo el valor de la función de supervivencia se obtiene que

$$q_i = 1 - e^{-\Lambda_o(a_i) + \Lambda_o(a_{i-1})}$$

o lo que es lo mismo,

$$1 - q_i = e^{-\Lambda_o(a_i) + \Lambda_o(a_{i-1})}, \qquad i=2,\ldots,k \tag{12}$$

y

$$1 - q_1 = e^{-\Lambda_o(a_1)} \tag{13}$$

Y a partir de (12) y (13), multiplicando se obtiene que

$$e^{-\Lambda_o(a_i)} = \prod_{j=1}^{i} (1 - q_j), \qquad i=1,2,\ldots,k$$

o lo que es lo mismo,

$$\Lambda_o(a_i) = \sum_{j=1}^{i} (-\log(1 - q_j)) = \sum_{j=1}^{i} r_j$$

en donde $r_j = -\log(1 - q_j)$, $j=1,2,\ldots,k$.

Especificar una distribución sobre el espacio de las trayectorias del proceso $S(t/\Lambda_o)$ es, por una parte, equivalente a especificar una distribución sobre el espacio de las trayectorias del proceso $\Lambda_o(t)$, y a la vista de la proposición acabada de demostrar, equivalente a especificar la distribución finito dimensional de las variables aleatorias $r_1,\ldots,r_k$ sobre cada partición $(a_{i-1}, a_i]$ $i=1,\ldots,k$

de forma que se cumplan ciertas condiciones de consistencia.

El problema pues, se reduce a la especificación de un proceso, que por la proposición tiene que ser de incrementos independientes no decreciente, y por otra parte las distribuciones de las variables aleatorias r_i, deben de estar sujetas a la condición de que la distribución de $r_i + r_{i+1}$ debe de ser la misma que la obtenida por aplicación directa de las reglas al intervalo combinado $(a_{i-1}, a_{i+1}]$

<u>Definición 1.6.1</u>. Sea $c > o$ y $\Lambda^*(t)$ una función de t, tal que $\exp[-\Lambda^*(t)]$ sea una función de supervivencia no aleatoria. Diremos que la función de supervivencia aleatoria $S(t/\Lambda_o)$ dada por (11), la función de distribución aleatoria y la función de probabilidad aleatoria asociadas, son un proceso gamma exponencial de parámetros $(\Lambda^*(t), c)$, si para cada partición finita $(a_{i-1}, a_i]$, $i=1,\ldots,k$, de $[0,\infty)$, las variables aleatorias $r_1,\ldots,r_k$ (los incrementos del proceso Λ_o) definidas anteriormente, se distribuyen independientemente como gammas de parámetros $(c(\Lambda^*(a_i) - \Lambda^*(a_{i-1})), c)$, $i=1,\ldots,k$. Es decir,

$$r_i \equiv G(c(\Lambda^*(a_i) - \Lambda^*(a_{i-1})), c)$$

Al parámetro $\Lambda^*(t)$ se le llama parámetro de forma y al parámetro c, de escala.

Vamos a ver a continuación otras dos definiciones de procesos gamma exponenciales que demostraremos más adelante, son equivalentes.

Definición 1.6.2. Sea $c > o$ y $\Lambda^*(t)$ una función de t, tal que $\exp\left[-\Lambda^*(t)\right]$ sea una función de supervivencia no aleatoria. Diremos que la función de distribución aleatoria $F(t/\Lambda_o)$, la función de supervivencia aleatoria y la función de probabilidad aleatoria asociadas, son un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$, si $F(t/\Lambda_o)$ es un proceso neutral a la derecha,

$$F(t/\Lambda_o) = 1 - e^{-\Lambda_o(t)}, \quad t \geq 0$$

en donde Λ_o es un proceso gamma de la forma, $\Lambda_o \sim G(c\Lambda^*,c)$; es decir, fijado un índice $t \geq 0$, $\Lambda_o(t)$ es una variable aleatoria gamma $G(c\ \Lambda^*(t),c)$.

Definición 1.6.3. Sea $c > o$ y $\Lambda^*(t)$ una función de t, tal que $\exp\left[-\Lambda^*(t)\right]$ sea una función de supervivencia no aleatoria. Diremos que la función de distribución aleatoria $F(t/\Lambda_o)$, la función de supervivencia aleatoria y la función de probabilidad aleatoria asociadas, son un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$, si $F(t/\Lambda_o)$ es un proceso neutral a la derecha, de la forma

$$F(t/\Lambda_o) = 1 - e^{-\Lambda_o(t)}, \quad t \geq 0$$

en donde el logaritmo de la función generatriz de momentos de $\Lambda_o(t)$ es de la forma

$$\log E\left[e^{-\theta\Lambda_o(t)}\right] = c\ \Lambda^*(t)\int_0^\infty (e^{-\theta z} - 1)dN(z)$$

siendo $N(.)$, la medida de Lévy asociada, de la forma

$$dN(z) = \frac{dz}{e^{cz} \cdot z}$$

En primer lugar hay que observar que la definición 1.6.3, es correcta, es decir, que la medida de Lévy asociada verifica

$$\int_0^\infty \frac{z}{1+z}\, dN(z) < \infty \qquad y \qquad \int_0^\infty z^2\, dN(z) < \infty$$

Proposición 1.6.2. Si $c > o$ y $N(.)$ es una medida sobre (o,∞) tal que

$$dN(z) = \frac{dz}{e^{cz} \,.\, z}$$

entonces,

$$\int_0^\infty \frac{z}{1+z}\, dN(z) < \infty \qquad y \qquad \int_0^\infty z^2\, dN(z) < \infty$$

En efecto:

$$\int_0^\infty \frac{z}{1+z}\, dN(z) = \int_0^\infty \frac{1}{(1+z)e^{cz}}\, dz = \int_0^{1/c} \frac{dz}{(1+z)e^{cz}} + \int_{1/c}^\infty \frac{dz}{(1+z)e^{cz}}$$

como $e^{cz} \geq 1 \Longrightarrow \frac{1}{e^{cz}} \leq 1$, con lo que

$$\leq \int_0^{1/c} \frac{dz}{1+z} + \int_{1/c}^\infty \frac{dz}{(1+z)e^{cz}}$$

si $z \geq \frac{1}{c} \Longrightarrow \frac{1}{1+z} \leq \frac{c}{c+1}$.

Por otro lado, como $e^{cz} = 1 + cz + \frac{c^2z^2}{2} + \frac{c^3z^3}{6} + \ldots \Longrightarrow$

$\Longrightarrow e^{cz} \geq \frac{c^2z^2}{2} \Longrightarrow \frac{1}{e^{cz}} \leq \frac{2}{c^2z^2}$

Así pues,

$$\int_0^\infty \frac{z}{1+z}\, dN(z) \leq \int_0^{1/c} \frac{dz}{1+z} + \int_{1/c}^\infty \frac{c}{c+1} \,.\, \frac{2}{c^2z^2}\, dz = \ln(1 + \frac{1}{c}) +$$

$$+ \frac{2}{c+1} < \infty$$

Por otro lado,

$$\int_0^\infty z^2 \, dN(z) = \int_0^\infty z \, e^{-cz} \, dz = \frac{1}{c^2} < \infty$$

con lo que la proposición queda demostrada.

Proposición 1.6.3. Sea una función $\gamma(t) > 0$ definida $\forall t \geq 0$ y acotada; sea $\tau \in R$ una constante positiva. Entonces,

$$\left(\frac{\tau}{\theta+\tau}\right)^{\gamma(t)} = e^{\gamma(t) \int_0^\infty (e^{-\theta z} - 1) e^{-\tau z} z^{-1} \, dz}$$

En efecto:

Consideremos un $t \geq 0$ fijo y a partir de los números $\gamma(t)$, τ, construyamos una variable aleatoria gamma Y_t con parámetros $(\gamma(t),\tau)$. La función generatriz de momentos (obsérvese el signo menos) será

$$M_t(\theta) = E\left[e^{-Y_t\theta}\right] = \int_0^\infty \frac{e^{-\theta y} \, e^{-\tau y} \, y^{\gamma(t)-1} \, \tau^{\gamma(t)}}{\Gamma(\gamma(t))} \, dy = \left(\frac{\tau}{\theta+\tau}\right)^{\gamma(t)}.$$

Por otra parte, la función característica de la variable artificial Y_t será

$$\phi_t(\theta) = \left(1 - \frac{i\theta}{\tau}\right)^{-\gamma(t)}$$

que es infinitamente divisible ya que existe una función, la

$$\phi_{t,n}(\theta) = \left(1 - \frac{i\theta}{\tau}\right)^{-\frac{\gamma(t)}{n}}$$

tal que

$$\left[\phi_{t,n}(\theta)\right]^n = \phi_t(\theta)$$

siendo ésta $\phi_{t,n}(\theta)$ la función característica correspondiente a una variable aleatoria gamma de parámetros $(\frac{\gamma(t)}{n}, \tau)$.

Así pues, $\phi_t(\theta)$ admite la representación de Lévy-Kolmogorov,

$$\log \phi_t(\theta) = i\,\alpha\,\theta + \int_0^\infty (e^{i\theta u} - 1 - i\theta u)\,\frac{1}{u^2}\,dK(u)$$

en donde

$$\alpha = \frac{\left[\frac{d \log \phi(t)}{dt}\right]_{t=o}}{i}$$

y $K(u)$ es una función que determinaremos a continuación.

Sea $K_n(x) = n \int_o^x z^2 f_n(z)dz$, siendo $f_n(z)$ la función de densidad de una variable aleatoria gamma $(\frac{\gamma(t)}{n}, \tau)$. Así pues

$$K_n(x) = n \frac{\tau^{\frac{\gamma(t)}{n}}}{\Gamma(\frac{\gamma(t)}{n})} \int_o^x z^2\, z^{\frac{\gamma(t)}{n} - 1}\, e^{-\tau z}\, dZ =$$

$$= a_n \int_o^x z^{\frac{\gamma(t)}{n} + 1}\, e^{-\tau z}\, dz$$

llamando $a_n = \dfrac{n\,\tau^{\frac{\gamma(t)}{n}}}{\Gamma(\frac{\gamma(t)}{n})} = \dfrac{\gamma(t)\,\tau^{\gamma(t)}}{\Gamma(1 + \frac{\gamma(t)}{n})}$. Se tiene que $\lim_{n\to\infty} a_n = \gamma(t)$

con lo que

$$K(x) = \lim_{n\to\infty} K_n(x) = \gamma(t) \int_o^x z\, e^{-\tau z}\, dz$$

y $K(+\infty) = \mathrm{Var}\,(Y_t) = \dfrac{\gamma(t)}{\tau^2} < \infty$

siendo además $\alpha = \dfrac{\gamma(t)}{\tau}$ con lo que

$$\log \phi_t(\theta) = i\,\frac{\gamma(t)}{\tau}\,\theta + \int_o^\infty (e^{i\theta u} - 1 - i\theta u)\,\frac{\gamma(t)\,u\,e^{-\tau u}}{u^2}\,du =$$

$$= \frac{i\gamma(t)\theta}{\tau} + \gamma(t) \int_0^\infty (e^{i\theta u} - 1)\, e^{-\tau u}\, u^{-1}\, du - i\theta\gamma(t) \int_0^\infty e^{-\tau u}\, du =$$

$$= \gamma(t) \int_0^\infty (e^{i\theta u} -1)\, e^{-\tau u}\, u^{-1}\, du,$$

y pasando a la función generatriz de momentos,

$$\log M_t(\theta) = \log \phi_t(i\theta) = \gamma(t) \int_0^\infty (e^{-\theta u} -1)\, e^{-\tau u}\, u^{-1}\, du$$

y dado que la representación de Lévy-Kolmogorov es única, deberá de ser

$$\left(\frac{c}{c+\theta}\right)^{\gamma(t)} = e^{\gamma(t) \int_0^\infty (e^{-\theta u} -1)\, e^{-\tau u}\, u^{-1}\, du}$$

como queríamos demostrar.

Proposición 1.6.4. Las definiciones 1.6.1, 1.6.2 y 1.6.3 son equivalentes.

En efecto:

El demostrar la equivalencia entre las definiciones 1.6.1 y 1.6.2 es sencillo. Como sabemos por la definición 1.5.1, una función de distribución aleatoria como $F(t/\Lambda_0)$, $t \geq 0$, es un proceso neutral a la derecha si y sólo si puede ser escrito en la forma

$$F(t/\Lambda_0) = 1 - e^{-\Lambda_0(t)}, \qquad t \geq 0$$

en donde Λ_0 es un proceso con incrementos independientes con ciertas propiedades. Pero decir que $F(t/\Lambda_0)$ es un proceso gamma exponencial equivale a decir que para toda partición finita de $[0,\infty)$ las variables aleatorias r_i (los incrementos del proceso Λ_0) se distribuyen independientemente como variables aleatorias gammas

$$r_i \equiv G(c(\Lambda^*(a_i) - \Lambda^*(a_{i-1})), c)$$

Consideremos, para t fijo, la partición $[0,t)$, $[t,\infty)$, será pues

$$\Lambda_o(t) \equiv G(c\ \Lambda^*(t), c)$$

por ser $\Lambda^*(0) = 0$ para que $\exp[-\Lambda^*(t)]$ sea función de supervivencia no aleatoria.

Por otro lado, si $\Lambda_o(t) \equiv G(c\ \Lambda^*(t), c)$ $\forall t$ fijo, entonces para toda partición de $[0,\infty)$ los incrementos (variables aleatorias independientes por ser Λ_o un proceso con incrementos independientes) serán

$$r_i = \Lambda_o(a_i) - \Lambda_o(a_{i-1})$$

con lo que

$$r_1 = \Lambda_o(a_1) \equiv G(c\ \Lambda^*(a_1), c)$$

$$r_2 = \Lambda_o(a_2) - \Lambda_o(a_1)$$

o lo que es lo mismo

$$r_1 + r_2 = \Lambda_o(a_1) + r_2 = \Lambda_o(a_2)$$

siendo r_1 y r_2 independientes, y $r_1 \equiv G(c\ \Lambda^*(a_1), c)$ y $\Lambda_o(a_2) \equiv$ $\equiv G(c\ \Lambda^*(a_2), c)$. Como r_1 y r_2 son independientes, serán las funciones características

$$\phi_{r_1}(t) \cdot \phi_{r_2}(t) = \phi_{\Lambda_o(a_2)}(t)$$

o lo que es lo mismo

$$(1 - \frac{it}{c})^{-c\Lambda^*(a_1)} \cdot \phi_{r_2}(t) = (1 - \frac{it}{c})^{-c\Lambda^*(a_2)}$$

con lo que

$$\phi_{r_2}(t) = (1 - \frac{it}{c})^{-c(\Lambda^*(a_2) - \Lambda^*(a_1))}$$

y por tanto, $r_2 \equiv G(c(\Lambda^*(a_2) - \Lambda^*(a_1)), c)$.

Análogamente, $\forall i=2,3,\ldots,k$

$$r_i = \Lambda_o(a_i) - \Lambda_o(a_{i-1}) = \Lambda_o(a_i) - r_1 - r_2 - \ldots - r_{i-1}$$

de donde

$$r_1 + r_2 + \ldots + r_{i-1} + r_i = \Lambda_o(a_i)$$

y por ser independientes las variables aleatorias $r_1, r_2, \ldots, r_i$, pasando a funciones características será

$$(1 - \frac{it}{c})^{c\Lambda^*(a_1)} (1 - \frac{it}{c})^{c\Lambda^*(a_2) - c\Lambda^*(a_1)} \cdot \ldots \cdot$$

$$(1 - \frac{it}{c})^{c\Lambda^*(a_{i-1}) - c\Lambda^*(a_{i-2})} \phi_{r_i}(t) = (1 - \frac{it}{c})^{c\Lambda^*(a_i)}$$

por ser $r_j \equiv G(c(\Lambda^*(a_j) - \Lambda^*(a_{j-1})), c)$, $j=1,2,\ldots,i-1$ y ser $\Lambda_o(a_i) = G(c\,\Lambda^*(a_i), c)$, con lo que la función característica de r_i será,

$$\phi_{r_i}(t) = (1 - \frac{it}{c})^{c\Lambda^*(a_i) - c\Lambda^*(a_{i-1})}$$

con lo que $r_i \equiv G(c(\Lambda^*(a_i) - \Lambda^*(a_{i-1})), c)$, $i=1,2,\ldots,k$, para toda partición finita $(a_{i-1}, a_i]$, $i=1,\ldots,k$ de $[0,\infty)$.

Así pues, las definiciones 1.6.1 y 1.6.2 son equivalentes.

Por otro lado, si Λ_o es un proceso gamma de parámetros $\Lambda^*(t)$, fijado $t \geq 0$, la variable aleatoria $\Lambda_o(t)$ es una variable aleatoria gamma de parámetros $(c\ \Lambda^*(t),\ c)$, con lo que de forma totalmente análoga a la realizada en la proposición 1.6.3, el logaritmo de la función generatriz de momentos de $\Lambda_o(t)$ se podrá escribir como,

$$\log M_{\Lambda_o(t)}(\theta) = c\ \Lambda^*(t) \int_0^\infty (e^{-\theta u} - 1)\ e^{-cu}\ u^{-1}\ du$$

o sea,

$$\log M_{\Lambda_o(t)}(\theta) = c\ \Lambda^*(t) \int_0^\infty (e^{-\theta u} - 1) d\ N(u)$$

siendo $N(.)$ la medida de Lévy asociada, que es de la forma

$$dN(u) = \frac{du}{e^{-cu}\ u}$$

Por último, si la medida de Lévy asociada es la dada arriba, el logaritmo de la función generatriz de momentos de $\Lambda_o(t)$ será de la forma

$$\log M_{\Lambda_o(t)}(\theta) = c\ \Lambda^*(t)\ .\int_0^\infty (e^{-\theta z} - 1) e^{-cz}\ z^{-1}\ dz$$

y como $c > o$ y $c\ \Lambda^*(t) > 0$ y acotada para que $\exp\left[-\Lambda^*(t)\right]$ sea función de supervivencia no aleatoria, aplicando la proposición 1.6.3, será

$$\log M_{\Lambda_o(t)}(\theta) = c\ \Lambda^*(t) \int_0^\infty (e^{-\theta z} - 1) e^{-cz}\ z^{-1}\ dz =$$

$$= \left(\frac{c}{c+\theta}\right)^{c\Lambda^*(t)} = \left(1 + \frac{\theta}{c}\right)^{-c\Lambda^*(t)}$$

que es la función generatriz de momentos de una variable aleatoria gamma de parámetros $(c\ \Lambda^*(t),\ c)$, y la proposición queda demostrada.

Conviene observar que en algunas ocasiones, por simplificar la notación, abreviamos $S(t/\Lambda_o)$ por $S(t)$ y $F(t/\Lambda_o)$ por $F(t)$ cuando no exista confusión posible sobre Λ_o.

Proposición 1.6.5. Sea $S(t/\Lambda_o)$ un proceso gamma exponencial de parámetros $(\Lambda^*(t),\ c)$. Entonces existe una medida de probabilidad sobre el espacio de las trayectorias del proceso $\Lambda_o(t)$ y equivalentemente, sobre el espacio de las trayectorias de $S(t/\Lambda_o)$; es decir, sobre el espacio paramétrico F.

En efecto:

Las $r_1,\ldots,r_k$ son variables aleatorias independientes y $r_i \equiv G(c(\Lambda^*(a_i) - \Lambda^*(a_{i-1})),\ c)$, $i=1,\ldots,k$ para cada partición $(a_{i-1},a_i]$, y como la variable aleatoria gamma es reproductiva con respecto al primer parámetro, la distribución de r_i+r_{i+1} será una gamma de parámetros, $(c(\Lambda^*(a_{i+1}) - \Lambda^*(a_i)),\ c)$, que es la misma que si considerásemos la correspondiente al intervalo $(a_{i-1},a_{i+1}]$; es decir, la suma de las distribuciones correspondientes a los intervalos $(a_{i-1},a_i]$ y $(a_i,a_{i+1}]$ $(a_{i-1},a_{i+1}]$, con lo cual se cumplen las condiciones de consistencia y por tanto existe una medida de probabilidad sobre el espacio de las trayectorias, como queríamos demostrar. A dicha medida la llamaremos "inducida" por el proceso gamma exponencial.

Como se ve, la definición 1.6.1 tiene ventajas con respecto a la 1.6.2 y a la 1.6.3 en el sentido de facilitar la demostración de la existencia de una medida de probabilidad sobre F y en el de manejar funciones típicas de un análisis de supervivencia como las funciones de azar q_i.

Vamos a ver a continuación unas propiedades de los procesos gamma exponenciales que más tarde, en la resolución de los problemas estadísticos que nos plantearemos, nos serán de mucha utilidad.

Media del proceso gamma exponencial:

Proposición 1.6.6. Sea $S(t/\Lambda_o)$ un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$, entonces su media es $(\frac{c}{c+1})^{\Lambda^*(t)c}$. Es decir,

$$E\left[S(t/\Lambda_o)\right] = E\left[e^{-\Lambda_o(t)}\right] = (\frac{c}{c+1})^{c\Lambda^*(t)}$$

en donde la esperanza es calculada con respecto a P, la probabilidad a priori.

En efecto:

Observemos que si consideramos la partición $([0,t],(t,\infty))$, tendríamos que

$$q = Pr\{T \in [0,t]\} = 1 - e^{-\Lambda_o(t)}$$

$$r = -\log(1-q) = \Lambda_o(t)$$

siguiendo $\Lambda_o(t)$ una distribución gamma de parámetros $(c.\Lambda^*(t),c)$, ya que $\Lambda^*(0) = 0$. Así pues,

$$E\left[\Lambda_o(t)\right] = \frac{c.\Lambda^*(t)}{c} = \Lambda^*(t), \quad y$$

$$V(\Lambda_o(t)) = \frac{\Lambda^*(t)}{c}$$

con lo cual

$$E\left[e^{-\Lambda_o(t)}\right] = \int_o^\infty e^{-y} \frac{1}{\Gamma(\Lambda^*(t).c)} c^{\Lambda^*(t).c} y^{\Lambda^*(t).c-1} e^{-yc} dy$$

$$= \left(\frac{c}{c+1}\right)^{\Lambda^*(t).c} = S_o(t)$$

en donde la última igualdad se entiende como notación.

<u>Proposición 1.6.7</u>. La función $S_o(t) = E\left[S(t)/\Lambda_o\right]$ es una función de supervivencia no aleatoria.

En efecto:

$S_o(t) = \left(\frac{c}{c+1}\right)^{\Lambda^*(t).c}$ y $\Lambda^*(t)$ es tal que $e^{-\Lambda^*(t)}$ tiene que ser una función de supervivencia, por lo que deberá de ser

(1) $\lim\limits_{t\to o} \Lambda^*(t) = 0$

(2) $\lim\limits_{t\to \infty} \Lambda^*(t) = \infty$

(3) $\Lambda^*(t)$ monótona no decreciente

(4) $\Lambda^*(t)$ contínua por la derecha

Tendremos pues que

(1) $\lim\limits_{t\to o} S_o(t) = \lim\limits_{t\to o} \left(\frac{c}{c+1}\right)^{\Lambda^*(t).c} = 1$

(2) $\lim\limits_{t\to \infty} S_o(t) = \lim\limits_{t\to \infty} \left(\frac{c}{c+1}\right)^{\Lambda^*(t).c} = 0$

(3) $S_o(t)$ es monótona no decreciente

(4) $S_o(t)$ es contínua por la derecha.

Y por la proposición 1.2.1, $S_o(t)$ será una función de supervivencia no aleatoria, como queríamos demostrar.

Aunque más tarde volveremos sobre ésto, al ser $S_o(t)$ una función de supervivencia y ser $S_o(t) = E[S(t)/\Lambda_o]$, en el contexto Bayesiano, puede considerarse y de hecho así lo haremos, como la función de supervivencia "a priori", nombre con el que la designaremos de ahora en adelante.

Dicha función de supervivencia a priori tiene asociada una función de distribución $F_o(t) = 1 - S_o(t)$, que es la distribución marginal de la muestra y a la que llamaremos de ahora en adelante función de distribución "a priori" marginal de la muestra, o simplemente función de distribución a priori.

Definición 1.6.4. Sea $c \in R$ un número tal que $c > o$. Definimos una función $l(c)$ que toma valores entre cero y uno como

$$l(c) = \frac{\log\left(\frac{c+2}{c+1}\right)}{\log\left(\frac{c+1}{c}\right)}$$

en donde log significa logaritmo neperiano, como a lo largo de todo este trabajo a menos que se especifique lo contrario.

Proposición 1.6.8. Propiedades de la función $l(c)$:

La función $l(c)$ acabada de definir verifica las siguientes propiedades:

(a) $0 < l(c) < 1$, siendo $c \in R^+$.

(b) $l(c)$ es monótona creciente en $c \in R^+$.

(c) $l(c) \longrightarrow 0 \iff c \longrightarrow 0$.

(d) $l(c) \longrightarrow 1 \iff c \longrightarrow \infty$

En efecto:

(a) Para $c \in R^+$, será

$$1 < \frac{c+2}{c+1} < \frac{c+1}{c}$$

con lo que tomando logaritmos será,

$$0 < \log \frac{c+2}{c+1} < \log \frac{c+1}{c}$$

o lo que es lo mismo,

$$0 < l(c) < 1$$

(b) Si llamamos $g(c) = \log \frac{c+2}{c+1}$ y $h(c) = \log \frac{c+1}{c}$ $\forall c \in R^+$, claramente $h(c) > 0$, siendo

$$l(c) = \frac{g(c)}{h(c)} \qquad \text{y} \qquad l'(c) = \frac{g'(c) - h'(c)l(c)}{h(c)}$$

Por otro lado, como $-\frac{1}{c(c+1)} < -\frac{1}{(c+1)(c+2)}$ será, $h'(c) < g'(c)$, y como $l(c) < 1$ por el apartado anterior, será $l(c)h'(c) < g'(c) \iff g'(c) - l(c)h'(c) > 0$ con lo que $l'(c) > 0$ $\forall c \in R^+$ y $l(c)$ es monótona creciente en $c \in R^+$.

(c)

$$\lim_{c \to o} l(c) = \frac{\log 2}{\lim_{c \to o} (-\log c)} = 0$$

Por otro lado, si $\lim_{c \to d} l(c) = 0$, con $d \in R^+$, $d \neq 0$, la función no sería creciente contradiciendo a (b).

(d)

$$\lim_{c \to \infty} l(c) = \lim_{c \to \infty} \frac{1}{1 + \frac{2}{c}} = 1$$

De la misma manera que antes, si $\lim_{c \to a} l(c) = 1$, con $a \in \mathbb{R}^+$ despuúes de a la función no sería creciente contradiciendo (b).

Así pues, la proposición queda demostrada.

Teorema 1.6.1. Sea $S(t/\Lambda_o)$ un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$ con función de supervivencia a priori $S_o(t)$. Entonces, si $x > y$, es

$$E[S(y).S(x)] = E[S(x).S(y)] = (\frac{c}{c+1})^{c\Lambda^*(x)} (\frac{c+1}{c+2})^{c\Lambda^*(y)} =$$

$$= S_o(x) \; . \; [S_o(x)]^{l(c)}$$

En efecto:

$$E[S(x).S(y)] = E\left[e^{-(\Lambda_o(x) + \Lambda_o(y))}\right]$$

Sabemos que por ser Λ_o un proceso con incrementos independientes, para $x > y$, los incrementos $\Lambda_o(y) - \Lambda_o(0) = \Lambda_o(y)$ y $\Lambda_o(x) - \Lambda_o(y)$ son variables aleatorias independientes y distribuidas según gammas $G(c\,\Lambda^*(y),c)$ y $G(c(\Lambda^*(x) - \Lambda^*(y)),c)$, respectivamente.

Llamando $Y_1 = \Lambda_o(x) - \Lambda_o(y)$

$Y_2 = \Lambda_o(y)$

se tiene que

$$\Lambda_o(x) + \Lambda_o(y) = Y_1 + 2Y_2$$

y si denotamos por $\alpha_1 = c(\Lambda^*(x) - \Lambda^*(y))$ y por $\alpha_2 = c\,\Lambda^*(x)$, la función de densidad conjunta del vector aleatorio (Y_1,Y_2) será:

$$f(y_1,y_2) = f_1(y_1)\cdot f_2(y_2) = \frac{c^{\alpha_1}\, e^{-cy_1}\, y_1^{\alpha_1-1}}{\Gamma(\alpha_1)} \cdot \frac{c^{\alpha_2}\, e^{-cy_2}\, y_2^{\alpha_2-1}}{\Gamma(\alpha_2)}$$

con lo que,

$$E\left[S(x).S(y)\right] = \int_0^{\infty}\int_0^{\infty} e^{-y_1} \frac{c^{\alpha_1}\, e^{-cy_1}\, y_1^{\alpha_1-1}}{\Gamma(\alpha_1)} \cdot \frac{c^{\alpha_2}\, e^{-cy_2}\, y_2^{\alpha_2-1}}{\Gamma(\alpha_2)}\, dy_1 dy_2 =$$

$$= \left(\frac{c}{c+1}\right)^{\alpha_1} \left(\frac{c}{c+2}\right)^{\alpha_2} = \left(\frac{c}{c+1}\right)^{c\Lambda^*(x)} \cdot \left(\frac{c+1}{c+2}\right)^{c\Lambda^*(y)}$$

Por otro lado, como

$$S_o(y) = \left(\frac{c}{c+1}\right)^{c\Lambda^*(y)} \Longrightarrow \Lambda^*(y) = \frac{\log S_o(y)}{c \log \frac{c}{c+1}}$$

y por ser

$$\log S_o(y) = \log_e S_o(y) = \log_e \frac{c+1}{c+2} \cdot \log_{\frac{c+1}{c+2}} S_o(y)$$

será

$$\left(\frac{c+1}{c+2}\right)^{c\Lambda^*(y)} = \left[\left(\frac{c+1}{c+2}\right)^{\log S_o(y)}\right]^{\frac{1}{\log \frac{c}{c+1}}} =$$

$$= \left[\left(\frac{c+1}{c+2}\right)^{\log_{\frac{c+1}{c+2}} S_o(y)}\right]^{\frac{\log_e\left(\frac{c+1}{c+2}\right)}{\log_e\left(\frac{c}{c+1}\right)}} = \left[S_o(y)\right]^{l(c)}$$

Así pues, si $x > y$ es

$$E\left[S(x) \;.\; S(y)\right] = S_o(x) \;.\; \left[S_o(y)\right]^{l(c)}$$

como queríamos demostrar.

Obsérvese que en el caso de que fuera $y > x$, consideraríamos las variables aleatorias $\Lambda_o(x)$ y $\Lambda_o(y) - \Lambda_o(x)$ y el razonamiento sería el mismo, obteniéndose al final que

$$E[S(y).S(x)] = E[S(x).S(y)] = S_o(y)[S_o(x)]^{1(c)}$$

Es decir, dados $x,y \in R$, con $x \neq y$ $E[S(y).S(x)] =$ $= E[S(x).S(y)]$ es el valor de S_o en el mayor de ellos, multiplicado por el valor de S_o en el menor, elevado a $1(c)$.

<u>Teorema 1.6.2</u>. Sea $S(t/\Lambda_o)$ un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$ con función de supervivencia a priori $S_o(t)$. Entonces,

$$E[(S(t))^2] = E[S^2(t)] = [S_o(t)]^{1(c)+1} = (\frac{c}{c+2})^{c\Lambda^*(t)}$$

En efecto:

$$E[S^2(t)] = E\left[e^{-2\Lambda_o(t)}\right] = \int_o^\infty e^{-2y} \frac{c^{\Lambda^*(t).c}}{\Gamma(\Lambda^*(t).c)} y^{\Lambda^*(t).c-1} e^{-yc} dy$$

por ser $\Lambda_o(t)$ una variable aleatoria gamma de parámetros $(c.\Lambda^*(t),c)$; así pues,

$$E[S^2(t)] = (\frac{c}{c+2})^{c\Lambda^*(t)} = (\frac{c+1}{c+2})^{c\Lambda^*(t)} . (\frac{c}{c+1})^{c\Lambda^*(t)} =$$

$$= [S_o(t)]^{1(c)} . S_o(t) = [S_o(t)]^{1(c)+1}$$

como queríamos demostrar.

Como advertimos en la nota al final del enunciado del teorema 1.3.8, no consideraremos el caso de ser $V(S(t)) = 0$, es decir, para el caso de los procesos gamma exponenciales, de ser $[S_o(t)]^2 =$ $= [S_o(t)]^{1(c)+1}$. Así pues, supondremos siempre $S_o(t) \neq 0$ y $S_o(t) \neq 1$.

Corolario 1.6.1. Sea $S(t/\Lambda_o)$ un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$ con función de supervivencia a priori $S_o(t)$. Entonces, dados $x,y \in R^+$ es

$$E[S(x).S(y)] = E[S(y).S(x)] = \begin{cases} S_o(x)[S_o(y)]^{l(c)} & \text{si} \quad y \leq x \\ \\ S_o(y)[S_o(x)]^{l(c)} & \text{si} \quad x < y \end{cases}$$

y $E[S(x)] = S_o(x)$, siendo $S_o(t) = (\frac{c}{c+1})^{\Lambda^*(t).c}$ y

$$l(c) = \frac{\log\,(c+2) - \log\,(c+1)}{\log\,(c+1) - \log\,c}\,.$$

La demostración de este corolario se deduce trivialmente de los teoremas 1.6.1 y 1.6.2 y de los proposición 1.6.6.

Como se ve, $E[S(y).S(x)]$ es una función simétrica respecto a la diagonal del primer cuadrante.

Función de covarianza del proceso gamma exponencial:

Proposición 1.6.9. La función de covarianza $r(x,y)$ de un proceso gamma exponencial $S(t/\Lambda_o)$ de parámetros $(\Lambda^*(t),c)$ es

$$r(x,y) = \begin{cases} (\frac{c}{c+1})^{c\Lambda^*(x)}\left[(\frac{c+1}{c+2})^{c\Lambda^*(y)} - (\frac{c}{c+1})^{c\Lambda^*(y)}\right] = S_o(x)([S_o(y)]^{l(c)} - S_o(y)) \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{si} \quad y \leq x \\ \\ (\frac{c}{c+1})^{c\Lambda^*(y)}\left[(\frac{c+1}{c+2})^{c\Lambda^*(x)} - (\frac{c}{c+1})^{c\Lambda^*(x)}\right] = S_o(y)([S_o(x)]^{l(c)} - S_o(x)) \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{si} \quad x < y \end{cases}$$

en donde $S_o(t)$ y $l(c)$ son los anteriormente definidos.

En efecto:

La función de covarianza vendrá definida para $x,y \in R^+$ como

$$r(x,y) = E\left[(S(x)-S_o(x))(S(y)-S_o(y))\right] = E\left[S(x).S(y)\right] - S_o(x).$$
$$. S_o(y)$$

Así pues, si $y \leq x$ será, por el corolario 1.6.1,

$$r(x,y) = S_o(x)\ \left[S_o(y)\right]^{1(c)} - S_o(x).S_o(y) = S_o(x)(\left[S_o(y)\right]^{1(c)} -$$

$$- S_o(y)) = (\frac{c}{c+1})^{c\Lambda^*(x)}((\frac{c+1}{c+2})^{c\Lambda^*(y)} - (\frac{c}{c+1})^{c\Lambda^*(y)})$$

y si $x < y$,

$$r(x,y) = S_o(y)\left[S_o(x)\right]^{1(c)} - S_o(x)S_o(y) = S_o(y)(\left[S_o(x)\right]^{1(c)} -$$

$$- S_o(x)) = (\frac{c}{c+1})^{c\Lambda^*(y)}\ ((\frac{c+1}{c+2})^{c\Lambda^*(x)} - (\frac{c}{c+1})^{c\Lambda^*(x)})$$

con lo cual queda demostrada la proposición.

<u>Continuidad en media cuadrática del proceso gamma exponencial</u>:

<u>Proposición 1.6.10</u>. Si la función $\Lambda^*(t)$ de un proceso gamma exponencial $S(t/\Lambda_o)$ de parámetros $(\Lambda^*(t),c)$ es continua en t, entonces dicho proceso $S(t/\Lambda_o)$ es continuo en media cuadrática.

En efecto:

Para que un proceso sea continuo en media cuadrática, es suficiente que la función de covarianza sea continua en la diagonal. Sabemos que $r(x,x)$ es,

$$r(x,x) = (\frac{c}{c+1})^{c\Lambda^*(x)}\ \left[(\frac{c+1}{c+2})^{c\Lambda^*(x)} - (\frac{c}{c+1})^{c\Lambda^*(x)}\right] =$$

$$= (\frac{c}{c+2})^{c\Lambda^*(x)} - (\frac{c}{c+1})^{2c\Lambda^*(x)}$$

función que es continua si $\Lambda^*(x)$ lo es. Así pues, la proposición ha quedado demostrada.

Derivabilidad en media cuadrática del proceso gamma exponencial:

Proposición 1.6.11. Si la función $\Lambda^*(t)$ de un proceso gamma exponencial $S(t/\Lambda_0)$ de parámetros $(\Lambda^*(t),c)$ tiene derivada contínua en t, el proceso $S(t/\Lambda_0)$ es derivable en media cuadrática.

En efecto:

Para que $S(t)$ sea derivable en media cuadrática es suficiente que $S(t)$ tenga derivada y que la derivada mixta de $r(x,y)$ exista y sea contínua. Como éstas dependen de la derivada primera de $\Lambda^*(t)$, la existencia y continuidad de la derivada de $\Lambda^*(t)$ implican la derivabilidad de $S(t)$.

Integración en media cuadrática del proceso gamma exponencial:

Consideremos un proceso $S(t/\Lambda_0)$ gamma exponencial de parámetros $(\Lambda^*(t),c)$ con la hipótesis de ser $\Lambda^*(t)$ una función continua en t, con lo que $S(t/\Lambda_0)$ será contínuo en media cuadrática y todos sus momentos existirán.

Sea $Y = \int_0^\infty S(t)dt$ la variable aleatoria la cual nos representa la integral de Lebesgue del proceso $S(t)$ y que ya hemos considerado en la sección 1.3.

Proposición 1.6.12. En las condiciones antes expuestas, $E[Y]$ es el momento de primer orden de $F_0(t)$, la función de distribución a prio ri.

En efecto:

$$E[Y] = \int_{F'} \int_0^\infty S(t)dt \, dP(S) \tag{14}$$

siendo F' el espacio de todas las funciones de supervivencia y $P(S)$ la medida de probabilidad inducida por el proceso $S(t)$ en dicho espacio.

Aplicando el teorema de Fubini en (14), será

$$E[Y] = \int_0^\infty \int_{F'} S(t)dP(S)dt = \int_0^\infty E[S(t)]dt = \int_0^\infty S_0(t)dt =$$
$$= \int_0^\infty (1 - F_0(t))dt = \int_0^\infty t \, dF_0(t) = \mu_0^1$$

en donde la última igualdad se entiende como notación.

Damos a continuación una generalización de la anterior proposición que nos será de mucha utilidad más adelante.

<u>Teorema 1.6.3</u>. Sea $S(t/\Lambda_0)$ un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$ con función de distribución a priori asociada a $F_0(t)$. Sea $g(t) : [0,\infty) \longrightarrow R$ una función medible de la variable aleatoria T asociada a $S(t)$, tal que

$$\int_F \left(\int_0^\infty |g(t)| \, dF(t)\right)dP(F) < \infty. \tag{15}$$

Entonces, $\int_F \left(\int_0^\infty g(t)dF(t)\right)dP(F)$ es la esperanza de $g(T)$ respecto de la distribución $F_0(t)$.

En efecto:

Por verificarse (15) se podrá aplicar el teorema de Fubini y tendremos que

$$\int_F \left(\int_0^\infty g(t)dF(t)\right)dP(F) = \int_0^\infty g(t)\left(\int_F dF(t)dP(F)\right) =$$

$$= \int_0^\infty g(t)dF_o(t) = E\left[g(T)/F_o\right]$$

por verificarse (9) al ser $F_o(t)$ el parámetro del proceso neutral por la derecha $S(t/\Lambda_o)$.

Como se ve, la proposición 1.6.12 es un caso particular de éste teorema tomando $g(T) = T$, por ser

$$\int_0^\infty t\ dF(t) = -\int_0^\infty t\ dS(t) = \int_0^\infty S(t)dt$$

después de integrar por partes.

<u>Teorema 1.6.4</u>. Sea $S(t/\Lambda_o)$ un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$ con función de supervivencia a priori asociada $S_o(t)$. Sea $g(t) : [0,\infty) \longrightarrow R$ una función medible de T, tal que

$$\int_{F'} \left(\int_0^\infty |S(t)g(t)|dt\right)dP(S) < \infty$$

entonces,

$$\int_{F'} \left(\int_0^\infty S(t)g(t)dt\right)dP(S) = \int_0^\infty g(t)S_o(t)dt$$

La demostración es inmediata, por ser

$$\int_{F'} \left(\int_0^\infty S(t)g(t)dt\right)dP(S) = \int_0^\infty g(t)\ \left(\int_{F'} S(t)dP(S)\right)dt$$

por el teorema 1.3.6, y ser $\int_{F'} S(t)dP(S) = S_o(t)$ por la proposición 1.6.6.

Teorema 1.6.5. Sea $S(t/\Lambda_0)$ un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$, siendo $\Lambda^*(t)$ una función contínua de t. Sea $g : [0,\infty) \longrightarrow R^+ \cup \{0\}$ una función medible no negativa tal que $\forall n \in N$ es

$$\int_0^n g^2(s)ds < \infty$$

Supongamos que

$$\int_{F'} \left(\int_0^\infty g(t)S(t)dt\right)\left(\int_0^\infty g(u)S(u)du\right)dP(S) < \infty$$

Entonces, es

$$\int_{F'} \left(\int_0^\infty g(t)S(t)dt\right)\left(\int_0^\infty g(u)S(u)du\right)dP(S) =$$

$$= \int_0^\infty g(t)\ S_0(t)\left(\int_0^t g(u)\left[S_0(u)\right]^{1(c)}du\right)dt +$$

$$+ \int_0^\infty g(t)\ \left[S_0(t)\right]^{1(c)}\ \left(\int_t^\infty g(u)S_0(u)du\right)dt$$

En donde $S_0(t)$ es la función de supervivencia a priori asociada y $1(c)$ la función definida en 1.6.4.

En efecto:

Al ser $\Lambda^*(t)$ una función continua de t, el proceso $S(t/\Lambda_0)$ es continuo en media cuadrática por la proposición 1.6.10, con lo que se puede aplicar el teorema 1.3.4 y será,

$$\int_{F'} \left(\int_0^\infty g(t)S(t)dt\right)\left(\int_0^\infty g(u)S(u)du\right)dP(S) =$$

$$= \int_0^\infty\int_0^\infty g(t)g(u)\ \left(\int_{F'} S(t)S(u)dP(S)\right)dtdu =$$

$$= \int_0^\infty\int_0^\infty g(t)g(u)\ E\left[S(t)S(u)\right]dtdu =$$

$$= \int_0^\infty (\int_0^t g(t)g(u)\ E[S(t).S(u)]du)dt\ +$$

$$+ \int_0^\infty (\int_t^\infty g(t)g(u)\ E[S(t).S(u)]du)dt =$$

$$= \int_0^\infty g(t)\ S_o(t)\ (\int_0^t g(u)\ [S_o(u)]^{1(c)}du)dt\ +$$

$$+ \int_0^\infty g(t)\ [S_o(t)]^{1(c)}\ (\int_t^\infty g(u)\ S_o(u)du)dt$$

a partir del corolario 1.6.1, con lo que el teorema queda demostrado.

<u>Teorema 1.6.6</u>. Sea $S(t/\Lambda_o)$ un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$, siendo $\Lambda^*(t)$ una función continua de t. Sean $g_1 : [0,\infty) \longrightarrow R^+ \cup \{0\}$ y $g_2 : [0,\infty) \longrightarrow R^+ \cup \{0\}$ dos funciones medibles no negativas tales que $\forall n \in N$ es

$$\int_0^n g_1(s)ds < \infty \qquad y \qquad \int_0^n g_2(s)ds < \infty.$$

Supongamos que

$$\int_{F'} (\int_0^\infty g_i(t)S(t)dt)(\int_0^\infty g_j(u)S(u)du)dP(S) < \infty \qquad \text{para} \quad \begin{array}{l} i=1,2, \\ j=1,2, \end{array}$$

Entonces, es

$$\int_{F'} (\int_0^\infty g_1(t)S(t)dt)(\int_0^\infty g_2(u)S(u)du)dP(S) =$$

$$= \int_0^\infty g_1(t)S_o(t)(\int_0^t g_2(u)\ [S_o(u)]^{1(c)}du)dt\ +$$

$$+ \int_0^\infty g_1(t)\ [S_o(t)]^{1(c)}(\int_t^\infty g_2(u)\ S_o(u)du)dt =$$

$$= \int_0^\infty g_2(u)\ S_o(u)(\int_0^u g_1(t)\ [S_o(t)]^{1(c)}dt)du\ +$$

$$+ \int_0^\infty g_2(u) \; [S_o(u)]^{1(c)} \; (\int_u^\infty g_1(t) \; S_o(t)dt)du$$

en donde $S_o(t)$ es la función de supervivencia a priori asociada y $1(c)$ la función definida en 1.6.4.

En efecto:

Al ser $\Lambda^*(t)$ una función continua de t, el proceso $S(t/\Lambda_o)$ es continuo en media cuadrática por la proposición 1.6.10, con lo que se puede aplicar el teorema 1.3.5 y será,

$$\int_{F'} (\int_0^\infty g_1(t)S(t)dt)(\int_0^\infty g_2(u)S(u)du)dP(S) =$$

$$= \int_0^\infty\int_0^\infty g_1(t)g_2(u)(\int_{F'} S(t)S(u)dP(S))dtdu$$

$$= \int_0^\infty\int_0^\infty g_1(t)g_2(u) \quad E[S(t).S(u)]dt \; du.$$

Por otro lado, a partir del corolario 1.6.1, será

$$\int_0^\infty\int_0^\infty g_1(t) \; g_2(u) \quad E[S(t).S(u)] = \int_0^\infty g_1(t)S_o(t)(\int_0^t g_2(u)[S_o(u)]^{1(c)}$$

$$du)dt + \int_0^\infty g_1(t) \; [S_o(t)]^{1(c)}(\int_t^\infty g_2(u) \; S_o(u)du)dt$$

y también,

$$\int_0^\infty\int_0^\infty g_1(t)g_2(u) \; E[S(t).S(u)] =$$

$$= \int_0^\infty g_2(u)S_o(u)(\int_0^u g_1(t) \; [S_o(t)]^{1(c)}dt)du \; +$$

$$+ \int_0^\infty g_2(u) \; [S_o(u)]^{1(c)} \; (\int_u^\infty g_1(t) \; S_o(t)dt)du$$

con lo que el resultado queda demostrado. "

Teorema 1.6.7. Sea $S(t/\Lambda_o)$ un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$, siendo $\Lambda^*(t)$ una función continua de t. Sean $h_1 : [0,\infty) \longrightarrow R$, y $h_2 : [0,\infty) \longrightarrow R$ dos funciones de la variable aleatoria T, tales que sus derivadas respectivas $g_1(x) = \dfrac{dh_1(x)}{dx}$ y $g_2(y) = \dfrac{dh_2(y)}{dy}$ son funciones medibles no negativas que verifican la condición

$$\forall n \in N, \qquad \int_o^n g_i(s)ds < \infty \quad , \qquad i=1,2$$

Supongamos que $c_1 = \lim_{x\to\infty} h_1(x)\, e^{-\Lambda_o(x)} - \lim_{x\to o} h_1(x) < \infty$ y que $c_2 = \lim_{x\to\infty} h_2(x)\, e^{-\Lambda_o(x)} - \lim_{x\to o} h_2(x) < \infty$.

Supongamos por último que

$$\int_{F'} (\int_o^\infty |g_i(t)S(t)|dt)dP(S) < \infty, \qquad i=1,2,$$

y que

$$\int_{F'} (\int_o^\infty g_i(t)S(t)dt)(\int_o^\infty g_j(u)S(u)du)dP(S) \qquad , \quad i=1,2, \quad j=i,2.$$

Entonces,

$$\int_{F'} (\int_o^\infty h_1(x)dF(x))(\int_o^\infty h_2(y)dF(y))dP(F) = c_1\cdot c_2 - c_1\int_o^\infty g_2(y)S_o(y)dy -$$

$$- c_2 \int_o^\infty g_1(x)S_o(x)dx + \int_o^\infty g_1(x)S_o(x)(\int_o^x g_2(y)\,[S_o(y)]^{1(c)}dy)dx +$$

$$+ \int_o^\infty g_1(x)\,[S_o(x)]^{1(c)} \cdot (\int_x^\infty g_2(y)\, S_o(y)dy)dx$$

en donde $S_o(t)$ es la función de supervivencia a priori asociada y $l(c)$ la función definida en 1.6.4.

En efecto:

Al ser $\Lambda^*(t)$ una función continua en t, el proceso $S(t/\Lambda_o)$ es continuo en media cuadrática por la proposición 1.6.10, con lo que se puede aplicar el teorema 1.3.7 y será,

$$\int_{F'} \left(\int_0^\infty h_1(x)dF(x)\right)\left(\int_0^\infty h_2(y)dF(y)\right)dP(F) = c_1 \cdot c_2 -$$

$$- c_1 \int_0^\infty g_2(y)E[S(y)]\,dy - c_2 \int_0^\infty g_1(x)\; E[S(x)]\,dx +$$

$$+ \int_0^\infty\int_0^\infty g_1(x)g_2(y)\; E[S(x).S(y)]\,dxdy =$$

$$= c_1 \cdot c_2 - c_1 \int_0^\infty g_2(y)\; S_o(y)dy - c_2 \int_0^\infty g_1(x)\; S_o(x) +$$

$$+ \int_0^\infty g_1(x)\; S_o(x)\left(\int_0^x g_2(y)\; [S_o(y)]^{1(c)}dy\right)dx +$$

$$+ \int_0^\infty g_1(x)\; [S_o(x)]^{1(c)}\; \left(\int_x^\infty g_2(y)\; S_o(y)dy\right)dx$$

a partir del corolario 1.6.1 y por ser $E[S(t)] = S_o(t)$ por la proposición 1.6.6.

Obsérvese que en todos estos teoremas, las expresiones podían haber venido dadas en función de c y $\Lambda^*(x)$, al ser $S_o(t) = (\frac{c}{c+1})^{c\Lambda^*(t)}$ por la proposición 1.6.6.

Como se ve por la proposición 1.6.5, los procesos gamma exponenciales considerados, de parámetros $(\Lambda^*(t),c)$, inducen una medida de probabilidad sobre el espacio de todas las medidas de probabilidad definidas en un espacio muestral dado, es decir, sobre el espacio paramétrico de los problemas de decisión no paramétrica. Dicha medida la utilizaremos más adelante como medida de probabilidad a priori al considerar los problemas de decisión desde un punto de vista Bayesiano. De éstas consideraciones deducimos la importancia de los procesos gamma exponenciales.

VII. PROCESOS HOMOGENEOS SIMPLES

Vamos a estudiar en esta sección otro proceso homogéneo, es decir otro proceso neutral a la derecha $F(t/Y_t)$, el cual nos va a definir una medida de probabilidad a priori sobre el espacio paramétrico de todas las distribuciones de probabilidad, y cuyo proceso con incrementos independientes correspondiente $Y_t = -\log(1-F(t))$ no tiene puntos fijos de discontinuidad, ni parte no aleatoria, siendo su medida de Lévy asociada independiente de t, es decir de la forma $N_t(.) = \gamma(t)N(.)$, con $\gamma(t)$ una función continua no decreciente tal que $\lim_{t\to -\infty} \gamma(t) = 0$ y $\lim_{t\to +\infty} \gamma(t) = +\infty$.

Sea $T \geq 0$ una variable aleatoria observable la cual nos indica el tiempo de fallo de un sistema. Consideremos la función de supervivencia aleatoria

$$S(t/Y_t) = 1 - F(t/Y_t) = e^{-Y_t}, \qquad t \geq 0$$

que nos indica

$$\Pr\{T > t/Y_t\} = e^{-Y_t} = S(t/Y_t)$$

en donde Y_t es un proceso estocástico que especificaremos a través de su medida de Lévy asociada. Obsérvese que ésta probabilidad es la probabilidad aleatoria, verificándose que

(1) $\lim_{t \to o} Y_t = 0$ c.s.,

(2) $\lim_{t \to \infty} Y_t = \infty$ c.s.,

(3) $Y_t - Y_s \geq 0$ c.s., para $t > s$

para que se cumplan las condiciones (2), (3), y (1) de la definición 1.3.3.

Definición 1.7.1. Sea $c > 0$ una constante y $\Lambda^*(t)$ una función de t, tal que $\exp\left[-\Lambda^*(t)\right]$ sea una función de supervivencia no aleatoria. Diremos que la función de distribución aleatoria $F(t/Y_t)$, la función de supervivencia aleatoria $S(t/Y_t) = 1 - F(t/Y_t)$ y la función de probabilidad aleatoria asociadas, son un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$, si $F(t/Y_t)$ es un proceso neutral a la derecha

$$F(t/Y_t) = 1 - e^{-Y_t} \quad , \quad t \geq 0$$

en donde el logaritmo de la función generatriz de momentos de Y_t es de la forma

$$\log E\left[e^{-\theta Y_t}\right] = c\ \Lambda^*(t) \int_0^\infty (e^{-\theta z} - 1)dN(z)$$

siendo la medida de Lévy asociada $N(.)$ de la forma

$$dN(z) = \frac{dz}{e^{cz}(1 - e^{-z})}$$

Vamos a asegurarnos en primer lugar de que la medida de Lévy así definida verifica que

$$\int_0^\infty \frac{z}{1+z}\, dN(z) < \infty \qquad y \qquad \int_0^\infty z^2\, dN(z) < \infty$$

<u>Proposición 1.7.1</u>. Si $c > 0$ es una constante y $N(.)$ es una medida sobre $(0,\infty)$ tal que

$$dN(z) = \frac{dz}{e^{cz}\,(1 - e^{-z})}$$

entonces,

$$\int_0^\infty \frac{z}{1+z}\, dN(z) < \infty \qquad \text{y} \qquad \int_0^\infty z^2\, dN(z) < \infty$$

En efecto:

$$\int_0^\infty z^2\, dN(z) = \int_0^\infty \frac{z^2}{e^{cz}(1-e^{-z})}\, dz = \int_0^\infty \frac{z^2\, e^z}{e^{cz}(e^z - 1)}\, dz$$

como $e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \ldots \Longrightarrow e^z - 1 \geq \frac{z^2}{2} \Longrightarrow \frac{1}{e^z - 1} \leq \frac{2}{z^2}$;

y además, como $e^{cz} = 1 + cz + \frac{c^2 z^2}{2} + \frac{c^3 z^3}{6} + \frac{c^4 z^4}{24} + \ldots \Longrightarrow e^{cz} \geq$ $\geq \frac{c^4 z^4}{24} \Longrightarrow \frac{1}{e^{cz}} \leq \frac{24}{c^4 z^4}$

Además, si $z < \frac{1}{c}$ es $e^z < e^{1/c}$ de donde

$$\frac{z^2\, e^z}{e^{cz}\,(e^z - 1)} \leq 2\, e^{1/c}\, e^{-cz}$$

y si $z \geq 1/c$ es $e^z \geq e^{1/c}$ de donde

$$\frac{1}{e^z} \leq \frac{1}{e^{1/c}} \Longrightarrow 1 - \frac{1}{e^z} \geq \frac{e^{1/c} - 1}{e^{1/c}}$$

y por tanto

$$\frac{z^2\, e^z}{e^{cz}(e^z - 1)} = \frac{z^2}{e^{cz}(1 - \frac{1}{e^z})} \leq \frac{z^2}{e^{cz}} \frac{e^{1/c}}{e^{1/c} - 1} \leq \frac{24 . z^2}{c^4 z^4} \cdot \frac{e^{1/c}}{e^{1/c} - 1} =$$

$$= \frac{24\, e^{1/c}}{c^4 (e^{1/c} - 1)} \cdot \frac{1}{z^2}$$

Así pues,

$$\int_0^\infty z^2\, dN(z) = \int_0^{1/c} \frac{z^2 e^z}{e^{cz}(e^z-1)}\, dz + \int_{1/c}^\infty \frac{z^2 e^z}{e^{cz}(e^z-1)}\, dz \leq$$

$$\leq 2\, e^{1/c} \int_0^{1/c} e^{-cz}\, dz + \frac{24\, e^{1/c}}{c^4(e^{1/c}-1)} \int_{1/c}^\infty \frac{dz}{z^2} = \frac{2}{c}\, e^{1/c}\, (1 - \frac{1}{e}) +$$

$$+ \frac{24\, e^{1/c}}{c^3(e^{1/c} - 1)} < \infty$$

Por otro lado,

$$\int_0^\infty \frac{z}{1+z}\, dN(z) = \int_0^\infty \frac{z}{1+z} \frac{dz}{e^{cz}(1 - e^{-z})} = \int_0^\infty \frac{z}{1+z} \frac{e^z}{e^{cz}(e^z - 1)}\, dz$$

Como $e^{cz} \geq 1 \Longrightarrow \frac{1}{e^{cz}} \leq 1$ y como $e^z - 1 \geq z \Longrightarrow \frac{1}{e^z-1} \leq \frac{1}{z}$.

Además, si $z \geq 1/c \Longrightarrow \frac{1}{1+z} \leq \frac{c}{c+1}$ y $\frac{e^z}{e^z-1} = \frac{1}{1 - \frac{1}{e^z}} \leq$

$\leq \frac{1}{1 - \frac{1}{e^{1/c}}} = \frac{e^{1/c}}{e^{1/c} - 1}$, siendo además $\frac{1}{e^{cz}} \leq \frac{6}{c^3 z^3}$, con lo que

$$\int_0^\infty \frac{z}{1+z}\, dN(z) = \int_0^{1/c} \frac{z}{1+z} \frac{e^z}{e^{cz}(e^z-1)}\, dz + \int_{1/c}^\infty \frac{z}{1+z} \frac{c^z}{e^{cz}(e^z-1)}\, dz \leq$$

$$\leq e^{1/c} \int_0^{1/c} \frac{z}{1+z} \cdot \frac{1}{z}\, dz + \int_{1/c}^\infty \frac{e^{1/c}}{e^{1/c} - 1} \cdot \frac{c}{c+1} \cdot \frac{6\, z}{c^3\, z^3}\, dz =$$

$$= e^{1/c} \ln(1 + \frac{1}{c}) + \frac{e^{1/c}}{e^{1/c} - 1}\ \frac{1}{c+1}\ \frac{6}{c} < \infty$$

y la proposición queda demostrada.

<u>Proposición 1.7.2</u>. Sea $S(t/Y_t)$ un proceso homogéneo simple de parámetros $(\Lambda^*(t), c)$. Entonces existe una medida de probabilidad sobre el espacio de las trayectorias del proceso Y_t y equivalentemente,

sobre el espacio de las trayectorias de $S(t/Y_t)$; es decir sobre el espacio paramétrico $\mathcal{F}$.

La proposición es evidente considerando que $F(t/Y_t) = 1 - S(t/Y_t)$ es un proceso neutral por la derecha.

Media del proceso homogéneo simple:

Proposición 1.7.3. Sea $S(t/Y_t)$ un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$. Entonces su media es $e^{-\Lambda^*(t)}$. Es decir,

$$E[S(t/Y_t)] = E\left[e^{-Y_t}\right] = e^{-\Lambda^*(t)}$$

en donde la esperanza es calculada con respecto a P, la probabilidad inducida por el proceso.

En efecto:

Por ser $S(t/Y_t)$ un proceso homogéneo simple, Y_t tiene por función generatriz de momentos, definición 1.7.1,

$$M_{Y_t}(\theta) = E\left[e^{-\theta Y_t}\right] = e^{c\Lambda^*(t)\int_0^\infty (e^{-\theta z}-1)\, e^{-cz}(1-e^{-z})^{-1}\, dz}$$

con lo que

$$E[S(t/Y_t)] = M_{Y_t}(1) = e^{-c\Lambda^*(t)\int_0^\infty e^{-cz} dz} = e^{-\Lambda^*(t)} = S_o(t)$$

en donde la última igualdad se entiende como notación.

Proposición 1.7.4. La función $S_o(t) = E[S(t)/Y_t] = e^{-\Lambda^*(t)}$ es una función de supervivencia no aleatoria.

La afirmación efectuada es cierta por el mero hecho de la condición impuesta a $\Lambda^*(t)$ en la definición 1.7.1.

Al igual que hicimos en el caso de los procesos gamma exponenciales, al ser $S_o(t) = E[S(t)/Y_t]$ una función de supervivencia no aleatoria, la consideraremos como la función de supervivencia "a priori", nombre con el que la designaremos de ahora en adelante, llamando a la correspondiente función de distribución $F_o(t) = 1 - S_o(t)$ función de distribución "a priori", que se corresponde con la distribución marginal de la muestra.

Definición 1.7.2. Sea $c > 0$ un número real. Definimos una función $m(c)$ que toma valores entre cero y uno como

$$m(c) = \frac{c}{c+1}$$

Proposición 1.7.5. La función $m(c)$ acabada de definir verifica las siguientes propiedades:

(a) $\forall c \in R^+, \quad 0 < m(c) < 1.$

(b) $m(c)$ es monótona creciente en $c \in R^+$.

(c) $m(c) \longrightarrow 0 \iff c \longrightarrow 0.$

(d) $m(c) \longrightarrow 1 \iff c \longrightarrow \infty.$

En efecto:

(a) Para $c \in R^+$ es

$$0 < c < c+1 \iff 0 < \frac{c}{c+1} < 1$$

o lo que es lo mismo,

$$0 < m(c) < 1$$

(b) $m'(c) = \frac{1}{(c+1)^2} > 0 \quad \forall c \in R^+$, con lo que la función es monótona creciente en $c \in R^+$.

(c) $\lim\limits_{c \to 0} m(c) = 0$

Por otro lado, si $m(c) = \frac{1}{1+\frac{1}{c}}$ tendiese hacia 0, debería de tender el denominar hacia $+\infty$, o equivalentemente c hacia cero.

(d) $\lim\limits_{c \to \infty} m(c) = \lim\limits_{c \to \infty} \frac{1}{1+\frac{1}{c}} = 1$

y de la misma manera, si $m(c) = \frac{c}{c+1}$ tiende hacia 1, ha de tender c hacia $+\infty$.

<u>Teorema 1.7.1</u>. Sea $S(t/Y_t)$ un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$, con función de supervivencia a priori $S_o(t)$. Entonces,

$$E\left[(S(t))^2\right] = E\left[S^2(t)\right] = \left[S_o(t)\right]^{m(c)+1}$$

En efecto:

$$E\left[S^2(t)\right] = E\left[e^{-2Y_t}\right] = M_{Y_t}(2) = e^{-c\Lambda^*(t)\int_0^\infty (e^{-z}+1)e^{-cz}dz} =$$

$$= e^{-\Lambda^*(t)\left[\frac{c}{c+1}+1\right]} = \left[S_o(t)\right]^{m(c)+1}$$

como queríamos demostrar.

<u>Teorema 1.7.2</u>. Sea $S(t/Y_t)$ un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$ con función de supervivencia a priori $S_o(t)$. Entonces, si $x > y$, es

$$E\left[S(x).S(y)\right] = E\left[S(y).S(x)\right] = e^{-\Lambda^*(x)} . e^{-\frac{c\Lambda^*(y)}{c+1}} =$$

$$= S_o(x) \cdot \left[S_o(y)\right]^{m(c)}$$

En efecto:

Por ser $S(t/Y_t)$ un proceso homogéneo simple, $F(t/Y_t) = 1 - S(t/Y_t)$ será un proceso neutral por la derecha y por tanto, Y_t un proceso con incrementos independientes.

Para $x > y$ consideremos los incrementos $Y_y - Y_o = Y_y$ y $Y_x - Y_y$ que serán variables aleatorias independientes.

Llamando

$$Z_1 = Y_x - Y_y$$

$$Z_2 = Y_y$$

se tiene que

$$Y_x + Y_y = Z_1 + 2Z_2$$

e

$$Y_x = Z_1 + Z_2$$

con lo que pasando a funciones características, será

$$\phi_{Y_x}(r) = \phi_{Z_1}(r) \cdot \phi_{Z_2}(r)$$

o bien,

$$\phi_{Z_1}(r) = \frac{\phi_{Y_x}(r)}{\phi_{Z_2}(r)} = \frac{\phi_{Y_x}(r)}{\phi_{Y_y}(r)}$$

y con funciones generatrices,

$$M_{Z_1}(\theta) = \frac{M_{Y_x}(\theta)}{M_{Y_y}(\theta)}$$

Por otro lado,

$$E[S(x).S(y)] = E\left[e^{-(Y_x+Y_y)}\right] = E\left[e^{-(Z_1+2Z_2)}\right] =$$

por ser Z_1 y Z_2 variables aleatorias independientes,

$$= E\left[e^{-Z_1}\right] . E\left[e^{-2Z_2}\right] = \frac{M_{Y_x}(1)}{M_{Y_y}(1)} . M_{Y_y}(2) =$$

$$= \frac{e^{-\Lambda^*(x)}}{e^{-\Lambda^*(y)}} . e^{-\Lambda^*(y)(\frac{c}{c+1}+1)} = e^{-\Lambda^*(x)} . e^{-\frac{c\Lambda^*(y)}{c+1}}$$

$$= S_o(x) . [S_o(y)]^{m(c)}$$

como queríamos demostrar.

Podríamos enunciar un nuevo teorema en el caso de ser $y > x$, aunque el razonamiento sería totalmente análogo considerando ahora las variables aleatorias Y_x e Y_y-Y_x y obteniendo que

$$E[S(y).S(x)] = E[S(x).S(y)] = e^{-\Lambda^*(y)} . e^{-\frac{c\Lambda^*(x)}{c+1}}$$

Podemos resumir por tanto los resultados de los teoremas 1.7.1 y 1.7.2 así como el de la proposición 1.7.3. en el siguiente corolario.

<u>Corolario 1.7.1</u>. Sea $S(t/Y_t)$ un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$ con función de supervivencia a priori $S_o(t)$. Entonces, dados $x,y \in R^+$ es

$$E[S(y).S(x)]=E[S(x).S(y)]=\begin{cases} S_o(x)[S_o(y)]^{m(c)} = e^{-(\Lambda^*(x)+\frac{c\Lambda^*(y)}{c+1})} & \text{si } y \leq x \\ \\ S_o(y)[S_o(x)]^{m(c)} = e^{-(\Lambda^*(y)+\frac{c\Lambda^*(x)}{c+1})} & \text{si } y > x \end{cases}$$

y $E[S(x)] = S_o(x)$, siendo $S_o(t) = e^{-\Lambda^*(t)}$ y $m(c) = \frac{c}{c+1}$.

Función de covarianza del proceso homogéneo simple:

$\forall x,y \in \mathbb{R}^+$ la función de covarianza $r(x,y)$ viene definida como

$$r(x,y) = E\left[(S(x)-S_o(x))(S(y)-S_o(y))\right] = E\left[S(x).S(y)\right] - $$
$$- S_o(x)S_o(y)$$

con lo que a partir del corolario 1.7.1. obtendremos inmediatamente el siguiente resultado.

Proposición 1.7.6. La función de covarianza $r(x,y)$ de un proceso homogéneo simple $S(t/Y_t)$ de parámetros $(\Lambda^*(t),c)$ y con función de supervivencia a priori $S_o(t)$ es

$$r(x,y) = \begin{cases} e^{-\Lambda^*(x)} \left(e^{-\frac{c\Lambda^*(y)}{c+1}} - e^{-\Lambda^*(y)}\right) = S_o(x)\left(\left[S_o(y)\right]^{m(c)} - S_o(y)\right) & \text{si } y \leq x \\ e^{-\Lambda^*(y)} \left(e^{-\frac{c\Lambda^*(x)}{c+1}} - e^{-\Lambda^*(x)}\right) = S_o(y)\left(\left[S_o(x)\right]^{m(c)} - S_o(x)\right) & \text{si } y > x \end{cases}$$

Continuidad en media cuadrática del proceso homogéneo simple:

Proposición 1.7.7. Si la función $\Lambda^*(t)$ de un proceso homogéneo simple $S(t/Y_t)$ de parámetros $(\Lambda^*(t),c)$ es continua en t, entonces dicho proceso $S(t/Y_t)$ es continuo en media cuadrática.

En efecto:

Deberemos comprobar si $r(x,y)$, la función de covarianza es continua en la diagonal

$$r(x,x) = e^{-\Lambda^*(x)} \left(e^{-\frac{c\Lambda^*(x)}{c+1}} - e^{-\Lambda^*(x)}\right) = e^{-\frac{2c+1}{c+1}\Lambda^*(x)} - e^{-2\Lambda^*(x)}$$

función que es continua si $\Lambda^*(x)$ lo es.

Derivabilidad en media cuadrática del proceso homogéneo simple:

Proposición 1.7.8. Si la función $\Lambda^*(t)$ de un proceso homogéneo simple $S(t/Y_t)$ de parámetros $(\Lambda^*(t),c)$ tiene derivada continua en t, el proceso $S(t/Y_t)$ es derivable en media cuadrática.

En efecto:

Para que $S(t)$ sea derivable en media cuadrática es suficiente que $S(t)$ tenga derivada y que la derivada mixta de $r(x,y)$ exista y sea continua. Como éstas dependen de la derivada primera de $\Lambda^*(t)$, la existencia y continuidad de la derivada de $\Lambda^*(t)$ implican la derivabilidad de $S(t)$.

Integración en media cuadrática del proceso homogéneo simple:

Sea $S(t/Y_t)$ un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$, y supongamos que $\Lambda^*(t)$ es una función continua de t, con lo que $S(t/Y_t)$ será continuo en media cuadrática y todos sus momentos existirán.

Sea $Y = \int_0^\infty S(t)dt$ la variable aleatoria la cual nos indica la integral de Lebesgue del proceso $S(t/Y_t)$.

Vamos a dar a continuación unos resultados que más tarde nos serán de mucha utilidad y que se deducen muy facilmente de los teoremas 1.3.4, 1.3.5, 1.3.6 y 1.3.7.

Teorema 1.7.3. Sea $S(t/Y_t)$ un proceso homogéneo simple de parámetros $(\Lambda^*/t),c)$ con función de distribución a priori asociada $F_0(t)$. Sea $g(t) : [0,\infty) \longrightarrow R$ una función medible de la variable aleatoria T asociada a $S(t/Y_t)$, tal que

$$\int_F \left(\int_0^\infty |g(t)|dF(t)\right)dP(F) < \infty \qquad (16)$$

Entonces, $\int_F \left(\int_0^\infty g(t)dF(t)\right)dP(F)$ es la esperanza de $g(T)$ con respecto a $F_o(t)$.

En efecto:

Por verificarse (16) se puede aplicar el teorema de Fubini, y al ser $F_o(t)$ el parámetro del proceso neutral por la derecha $S(t/Y_t)$ se obtiene el resultado,

$$\int_F \left(\int_0^\infty g(t)dF(t)\right)dP(F) = \int_0^\infty g(t)\left(\int_F dF(t)dP(F)\right) = \int_0^\infty g(t)dF_o(t)$$

<u>Corolario 1.7.2</u>. Sea $S(t/Y_t)$ un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$ con función de distribución a priori asociada $F_o(t)$. Entonces, $E[Y]$ es el momento de primer orden de $F_o(t)$.

Este resultado se obtiene sin más que tomar $g(T) = T$ en el teorema anterior.

<u>Teorema 1.7.4</u>. Sea $S(t/Y_t)$ un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$, con función de supervivencia a priori asociada $S_o(t)$. Sea $g(t) : [0,\infty) \longrightarrow R$ una función medible de T, tal que

$$\int_{F'} \left(\int_0^\infty |S(t)g(t)|dt\right)dP(S) < \infty$$

entonces

$$\int_{F'} \left(\int_0^\infty S(t)g(t)dt\right)dP(S) = \int_0^\infty g(t)S_o(t)dt$$

La demostración es inmediata utilizando primero el teorema 1.3.6 y después la proposición 1.7.3.

Teorema 1.7.5. Sea $S(t/Y_t)$ un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$, siendo $\Lambda^*(t)$ una función continua de t. Sea $g : [0,\infty) \longrightarrow R^+ \cup \{0\}$ una función medible no negativa tal que $\forall n \in N$ es

$$\int_0^n g^2(s)ds < \infty$$

Supongamos que

$$\int_{F'} (\int_0^\infty g(t)S(t)dt)(\int_0^\infty g(u)S(u)du)dP(S) < \infty$$

Entonces es

$$\int_{F'} (\int_0^\infty g(t)S(t)dt)(\int_0^\infty g(u)S(u)du)dP(S) =$$

$$= \int_0^\infty g(t)S_o(t)(\int_0^t g(u)[S_o(u)]^{m(c)}du)dt +$$

$$+ \int_0^\infty g(t)[S_o(t)]^{m(c)} (\int_t^\infty g(u)S_o(u)du)dt$$

en donde $S_o(t)$ es la función de supervivencia a priori asociada y $m(c)$ la función definida en 1.7.2.

En efecto:

Como $\Lambda^*(t)$ es una función continua, $S(t/Y_t)$ será continuo en media cuadrática, con lo que se podrá aplicar el teorema 1.3.4 y será

$$\int_{F'}(\int_0^\infty g(t)S(t)dt)(\int_0^\infty g(u)S(u)du)dP(S) =$$

$$= \int_0^\infty\int_0^\infty g(t)g(u)\ E[S(t)S(u)]dt\ du =$$

$$= \int_0^\infty g(t)S_o(t)(\int_0^t g(u)[S_o(u)]^{m(c)}\ du)dt +$$

$$+ \int_0^\infty g(t)[S_o(t)]^{m(c)}\ (\int_t^\infty g(u)S_o(u)du)dt$$

$$= \int_0^\infty g(t)\ e^{-\Lambda^*(t)}\ (\int_0^t g(u)\ e^{-\frac{c\Lambda^*(u)}{c+1}}\ du)dt +$$

$$+ \int_0^\infty g(t)\ e^{-\frac{c\Lambda^*(t)}{c+1}}\ (\int_t^\infty g(u)\ e^{-\Lambda^*(u)}\ du)dt$$

<u>Teorema 1.7.6</u>. Sea $S(t/Y_t)$ un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$, siendo $\Lambda^*(t)$ una función continua de t. Sean $g_1 : [0,\infty) \longrightarrow R^+ \cup \{0\}$ y $g_2 : [0,\infty) \longrightarrow R^+ \cup \{0\}$ dos funciones medibles no negativas tales que $\forall n \in N$ es

$$\int_0^n g_1(s)ds < \infty \qquad \text{y} \qquad \int_0^n g_2(s)ds < \infty$$

Supongamos que

$$\int_{F'} (\int_0^\infty g_i(t)S(t)dt)(\int_0^\infty g_j(u)S(u)du)dP(S) < \infty \qquad \text{para } i=1,2,\ j=i,2.$$

Entonces, es

$$\int_{F'} (\int_0^\infty g_1(t)S(t)dt)(\int_0^\infty g_2(u)S(u)du)dP(S) =$$

$$= \int_0^\infty g_1(t)S_o(t)(\int_0^t g_2[S_o(u)]^{m(c)}du)dt +$$

$$+ \int_0^\infty g_1(t)[S_o(t)]^{m(c)}\ (\int_t^\infty g_2(u)S_o(u)du)dt =$$

$$= \int_0^\infty g_2(u) S_o(u) \left(\int_0^u g_1(t) \left[S_o(t)\right]^{m(c)} dt \right) du +$$

$$+ \int_0^\infty g_2(u) \left[S_o(u)\right]^{m(c)} \left(\int_u^\infty g_1(t)\, S_o(t) dt \right) du$$

en donde $S_o(t)$ es la función de supervivencia a priori asociada y $m(c)$ la función definida en 1.7.2.

En efecto:

Dado que $\Lambda^*(t)$ es una función continua en t, el proceso $S(t/Y_t)$ es continuo en media cuadrática y por tanto se puede aplicar el teorema 1.3.5. El resultado se concluye a partir del corolario 1.7.1.

Obsérvese que en función de $\Lambda^*(t)$ y c el resultado quedaría de la forma

$$\int_{F'} \left(\int_0^\infty g_1(t) S(t) dt \right) \left(\int_0^\infty g_2(u) S(u) du \right) dP(S) =$$

$$= \int_0^\infty g_1(t)\, e^{-\Lambda^*(t)} \left(\int_0^t g_2(u)\, e^{-\frac{c\Lambda^*(u)}{c+1}} du \right) dt +$$

$$+ \int_0^\infty g_1(t)\, e^{-\frac{c\Lambda^*(t)}{c+1}} \left(\int_t^\infty g_2(u)\, e^{-\Lambda^*(u)} du \right) dt =$$

$$= \int_0^\infty g_2(u)\, e^{-\Lambda^*(u)} \left(\int_0^u g_1(t)\, e^{-\frac{c\Lambda^*(t)}{c+1}} dt \right) du +$$

$$+ \int_0^\infty g_2(u)\, e^{-\frac{c\Lambda^*(u)}{c+1}} \left(\int_u^\infty g_1(t)\, e^{-\Lambda^*(t)} dt \right) du$$

<u>Teorema 1.7.7</u>. Sea $S(t/Y_t)$ un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$, siendo $\Lambda^*(t)$ una función continua en t. Sean

$h_1 : [0,\infty) \longrightarrow R$ y $h_2 : [0,\infty) \longrightarrow R$ dos funciones de la variable aleatoria T, tales que sus derivadas respectivas $g_1(x) = \frac{dh_1(x)}{dx}$ y $g_2(y) = \frac{dh_2(y)}{dy}$ son funciones medibles no negativas que verifican la condición

$$\forall n \in N, \quad \int_0^n g_i(s)ds < \infty \quad , \quad i=1,2$$

Supongamos que $c_1 = \lim_{x\to\infty} h_1(x)\, e^{-\gamma_x} - \lim_{x\to 0} h_1(x) < \infty$ y que $c_2 = \lim_{x\to\infty} h_2(x)\, e^{-\gamma_x} - \lim_{x\to 0} h_2(x) < \infty$.

Supongamos por último que

$$\int_{F'} (\int_0^\infty |g_i(t)\ S(t)|dt)dP(S) < \infty \ , \quad i=1,2,$$

y que

$$\int_{F'} (\int_0^\infty g_i(t)S(t)dt)(\int_0^\infty g_j(u)S(u)du)dP(S) < \infty, \quad \begin{matrix} i=1,2, \\ j=i,2. \end{matrix}$$

Entonces,

$$\int_F (\int_0^\infty h_1(x)dF(x))(\int_0^\infty h_2(y)dF(y))dP(F) =$$

$$= c_1 \cdot c_2 - c_1 \int_0^\infty g_2(y)S_o(y)dy - c_2 \int_0^\infty g_1(x)S_o(x)dx +$$

$$+ \int_0^\infty g_1(x)S_o(x)(\int_0^x g_2(y)\ [S_o(y)]^{m(c)}\ dy)dx$$

$$+ \int_0^\infty g_1(x)\ [S_o(x)]^{m(c)} \ . \ (\int_x^\infty g_2(y)\ S_o(y)dy)dx =$$

$$= c_1 \cdot c_2 - c_1 \int_0^\infty g_2(y)\ e^{-\Lambda^*(y)}\ dy - c_2 \int_0^\infty g_1(x)\ e^{-\Lambda^*(x)}\ dx +$$

$$+ \int_0^\infty g_1(x)\ e^{-\Lambda^*(x)}\ (\int_0^x g_2(y)\ e^{-\frac{c\Lambda^*(y)}{c+1}}\ dy)dx \qquad "$$

$$+ \int_0^\infty g_1(x)\, e^{-\frac{c\Lambda^*(x)}{c+1}} \cdot \left(\int_x^\infty g_2(y)\, e^{-\Lambda^*(y)} dy\right) dx,$$

en donde $S_o(t)$ es la función de supervivencia a priori asociada y $m(x)$ la función definida en 1.7.2.

En efecto:

Al ser $\Lambda^*(t)$ una función continua en t, el proceso $S(t/Y_t)$ es continuo en media cuadrática por la proposición 1.7.7, con lo que se puede aplicar el teorema 1.3.7 y será, a partir del corolario 1.7.1,

$$\int_F \left(\int_0^\infty h_1(x)dF(x)\right)\left(\int_0^\infty h_2(y)dF(y)\right)dP(F) = c_1 \cdot c_2$$

$$- c_1 \int_0^\infty g_2(y)S_o(y)dy - c_2 \int_0^\infty g_1(x)S_o(x) +$$

$$+ \int_0^\infty g_1(x)S_o(x)\left(\int_0^x g_2(y)\ [S_o(y)]^{m(c)}\ dy\right)dx +$$

$$+ \int_0^\infty g_1(x)\ [S_o(x)]^{m(c)}\ \left(\int_x^\infty g_2(y)\ S_o(y)dy\right)dx$$

obteniendo la última igualdad por ser $S_o(t) = e^{-\Lambda^*(t)}$ y $[S_o(t)]^{m(c)} = e^{-\frac{c\Lambda^*(t)}{c+1}}$.

VIII. <u>PROCESOS GAMMA EXTENDIDOS</u>

Sea $T \geq 0$ una variable aleatoria observable, la cual nos indica el tiempo de fallo de un sistema.

<u>Definición 1.8.1</u>. (Dykstra y Purushottam Laud (1981): Sea $\alpha(t)$, $t \geq 0$ una función continua, real valorada, no decreciente con

$\alpha(0) = 0$. Sea $\beta(t)$, $t \geq 0$ una función continua, real valorada, y positiva, acotada siempre por cero y con límites a la izquierda existentes.

Sea $Z(t)$, $t \geq 0$, un proceso gamma con incrementos independientes correspondiente a $\alpha(t)$, definido sobre un apropiado espacio probabilístico $(\Omega, \mathcal{A}, P)$; es decir, $Z(0) \equiv 0$, $Z(t)$ tiene incrementos independientes, y para $t > s$, $Z(t) - Z(s)$ es una distribución Gamma de parámetros $(\alpha(t) - \alpha(s), 1)$. Supondremos por último que dicho proceso tiene trayectorias continuas por la derecha no decrecientes.

Llamaremos proceso gamma extendido al proceso

$$r(t) = \int_{[0,t)} \beta(s)\, dZ(s)$$

en donde la integración se realiza con respecto a las trayectorias del proceso $Z(t)$, y lo representaremos por $r(t) \equiv \Gamma(\alpha(.), \beta(.))$, suponiendo que dicha integral estocástica esté bien definida.

Como se ve, el proceso gamma extendido $r(t)$ es una tasa de azar aleatoria, con lo que

$$H(t) = \int_{[0,t)} r(s) ds$$

es una función de azar acumulativa aleatoria (la correspondiente a $r(t)$), y

$$S(t) = e^{-H(t)} = e^{-\int_{[0,t)} r(s) ds}$$

una función de supervivencia aleatoria, la correspondiente a $r(t)$, por lo que la representaremos por $S(t/r(t))$, en donde las integrales en general serán de Lebesgue.

Según el trabajo de Doksum (1974), $S(t/r(t))$ sería neutral a la derecha si $H(t)$ tuviera incrementos independientes, pero aunque $r(t)$ los tiene, $H(t) = \int_{[0,t)} r(s)ds$ no los tiene y por tanto no se puede considerar aplicable el trabajo de Doksum.

Proposición 1.8.1. (Dykstra y Laud (1981)): El proceso gamma extendido $r(t) \equiv \Gamma(\alpha(.), \beta(.))$ tiene incrementos independientes.

En efecto:

Se la sucesión de particiones $0 = t_{n,1} < t_{n,2} < \ldots < t_{n,k(n)}$ tal que $K(n) \xrightarrow[n \to \infty]{} \infty$ y $t_{n,k(n)} \xrightarrow[n \to \infty]{} t$ y cuya norma tiende a cero para $n \to \infty$.

La sucesión de funciones aleatorias, definidas para $t \geq 0$

$$r_n(t) = \begin{cases} \sum_{\{i>o;\ t_{n,i}<t\}} \beta(t_{n,i}) \left[Z(t_{n,i}) - Z(t_{n,i-1})\right] & \text{si } t > t_{n,2} \\ 0 & \text{si } t_{n,1} \leq t \leq t_{n,2} \end{cases}$$

es tal que para todo t fijo $\lim_{n \to \infty} r_n(t) = r(t)$, c.s., con lo que la proposición queda demostrada.

Proposición 1.8.2. El proceso gamma extendido $r(t) \equiv \Gamma(\alpha(.), \beta(.))$ induce una medida de probabilidad sobre el espacio de las trayectorias y consecuentemente sobre el espacio paramétrico F.

La demostración es evidente al ser $r(t)$ un proceso con incrementos independientes por la proposición 1.8.1. A dicha medida de probabilidad inducida la llamaremos "a priori" y la representaremos por P.

Hay que notar, como anteriormente hicimos, que siempre podremos encontrar la versión separable de $r(t)$ de tal manera que sus trayectorias sean funciones tasas de azar con probabilidad 1, y justamente a esa versión separable será a la que nos referiremmos en todo éste trabajo, al igual que cuando hablemos de funciones de distribución aleatorias o de supervivencias aleatorias.

Proposición 1.8.3. Sea $r(t) \equiv \Gamma(\alpha(.), \beta(.))$ un proceso gamma extendido. Entonces,

$$E[r(t)] = \int_{[0,t)} \beta(s)d\alpha(s)$$

y

$$V(r(t)) = \int_{[0,t)} \beta^2(s)d\alpha(s)$$

en donde dichas esperanza y varianza están calculadas con respecto a la distribución a priori P.

En efecto:

$$E[r(t)] = \lim_{n\to\infty} E[r_n(t)] = \lim_{n\to\infty} \sum_{\{i>o,t_{n,i}<t\}} \beta(t_{n,i})[\alpha(t_{n,i}) - \alpha(t_{n,i-1})] = \int_{[0,t)} \beta(s)d\alpha(s)$$

y

$$V(r(t)) = \lim_{n\to\infty} V(r_n(t)) = \lim_{n\to\infty} \sum_{\{i>o,t_{n,i}<t\}} \beta^2(t_{n,i}) [\alpha(t_{n,i}) - \alpha(t_{n,i-1})] = \int_{[0,t)} \beta^2(s)d\alpha(s)$$

Definición 1.8.2. Sea $\alpha(t)$, $t \geq 0$ una función continua, real valorada, no decreciente con $\alpha(0) = 0$. Sea $\beta(t)$, $t \geq 0$ una función continua, real valorada y positiva, acotada siempre por cero y con lí" mites a la izquierda existentes.

Diremos que la función de supervivencia aleatoria $S(t/r(t))$ es un proceso gamma extendido de parámetros $(\alpha(.), \beta(.))$ y lo representaremos por $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$ cuando la función tasa de azar $r(t)$ correspondiente a $S(t/r(t))$ es un proceso gamma extendido de los definidos en 1.8.1.

Dado que

$$1 - F(t/r(t)) = S(t/r(t)) = e^{-H(t)} = e^{-\int_{[0,t)} r(s)ds}$$

cuando nos refiramos a un proceso gamma extendido de parámetros $(\alpha(.), \beta(.))$ nos podremos referir a cualquiera de las funciones aleatorias asociadas, la función de distribución aleatoria $F(t/r(t))$, la función de supervivencia aleatoria $S(t/r(t))$, la función de azar acumulativa aleatoria $H(t)$, o la función tasa de azar aleatoria $r(t)$, siendo ésta última la que nos ha servido para introducir a los procesos gammas extendidos.

<u>Medida de la función de supervivencia aleatoria $S(t/r(t))$</u>:

<u>Proposición 1.8.4</u>. Sea $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$ un proceso gamma extendido; entonces,

$$E[S(t/r(t))] = S_o(t) = Pr\{T > t\} = e^{-\int_{[0,t)} \log(1+\beta(s)(t-s))d\alpha(s)}$$

en donde la primera igualdad se entiende como notación.

En efecto:

$$E[S(t/r(t))] = E\left[e^{-H(t)}\right] = E\left[e^{-\int_{[0,t)} r(s)ds}\right]$$

en donde

$$r(s) = \int_{[0,s)} r(x)\,dZ(x)$$

Consideremos la sucesión de particiones antes estudiada en la proposición 1.8.1, y por simplificar utilizaremos la siguiente notación: $t_i = t_{n,i}$ y $\Delta\alpha_i = \alpha(t_i) - \alpha(t_{i-1})$, $i=1,2,\ldots,k(n)$.

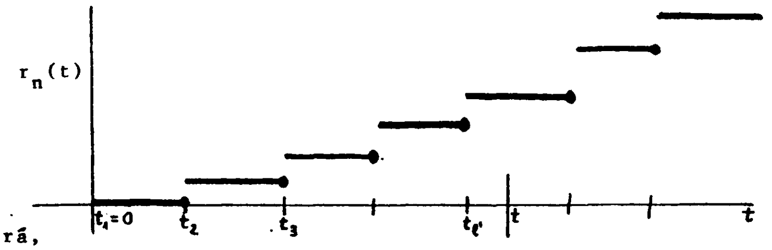
Con ésta nueva notación, $r_n(t)$ será ahora

$$r_n(t) = \begin{cases} \sum_{\{i>o;\,t_i<t\}} \beta_i\left[Z(t_i)-Z(t_{i-1})\right] & \text{si} \quad t > t_2 \\ 0 & \text{si} \quad t_1 \leq t \leq t_2 \end{cases}$$

Así pues,

$$E\left[S(t)/r(t)\right] = E\left[e^{-\int_{[0,t)} r(s)ds}\right] = \lim_{n\to\infty} E\left[e^{-\int_{[0,t)} r_n(s)ds}\right]$$

Por otro lado, si $t_{l'}$ es el mayor t_i menor que t,



será,

$$\int_{[0,t)} r_n(s)ds = \sum_{i=2}^{l'} r_n(t_i)\,(t_i-t_{i-1}) + r_n(t-t_{l'}) =$$

$$= 0.(t_2-t_1) + \Big(\beta(t_2)\ \left[Z(t_2)-Z(t_1)\right](t_3-t_2)\Big) +$$

$$+ \Big(\beta(t_2)\ \left[Z(t_2)-Z(t_1)\right]\ (t_4-t_3) + \beta(t_3)\ \left[Z(t_3)-Z(t_2)\right]\ (t_4-t_3)\Big) +$$

$$+ \ldots + \Big(\beta(t_2)\,[Z(t_2)-Z(t_1)]\,(t_{i'}-t_{i'-1}) + \beta(t_3)\,[Z(t_3)-Z(t_2)]$$

$$(t_{i'}-t_{i'-1}) + \ldots + \beta(t_{i'-1})\,[Z(t_{i'-1})-Z(t_{i'-1})]\,(t_{i'}-t_{i'-1})\Big) +$$

$$+ \Big(\beta(t_2)\,[Z(t_2)-Z(t_1)]\,(t-t_{i'}) + \beta(t_3)\,[Z(t_3)-Z(t_2)]\,(t-t_{i'}) +$$

$$+ \ldots + \beta(t_{i'-1})\,[Z(t_{i'-1})-Z(t_{i'-2})]\,(t-t_{i'}) +$$

$$+ \beta(t_{i'})\,[Z(t_{i'})-Z(t_{i'-1})]\,(t-t_{i'})\Big) = \sum_{i=2}^{i'} \beta(t_i)\,[Z(t_i)-Z(t_{i-1})](t-t_i)$$

con lo que

$$E[S(t)/r(t)] = \lim_{n\to\infty} E\left[e^{-\sum_{i=2}^{i'} \beta(t_i)[Z(t_i)-Z(t_{i-1})](t-t_i)}\right] =$$

$$= \lim_{n\to\infty} \prod_{i=2}^{i'} E\left[e^{-\beta(t_i)[Z(t_i)\,-\,Z(t_{i-1})](t-t_i)}\right]$$

por ser $Z(t_i)-Z(t_{i-1})$ variables aleatorias gammas de parámetros $(\alpha(t_i)-\alpha(t_{i-1}),1)$ e independientes, $i=2,3,\ldots,$ y se puede interpretar el último resultado como la función característica de una variable gamma en el punto $i\beta(t_i)\cdot(t-t_i)$, con lo que

$$E[S(t)/r(t)] = \lim_{n\to\infty} \prod_{i=2}^{i'} \left(\frac{1}{1+\beta(t_i)(t-t_i)}\right)^{\Delta\alpha_i} =$$

$$= \lim_{n\to\infty} \frac{1}{\prod_{i=2}^{i'} e^{\Delta\alpha_i \log(1+\beta(t_i)(t-t_i))}} =$$

$$= \frac{1}{\lim_{n\to\infty} e^{\sum_{i=2}^{i'} \Delta\alpha_i \log[1+\beta(t_i)\cdot(t-t_i)]}} =$$

$$= e^{-\int_{[0,t)} \log\left[1+\beta(s)(t-s)\right] d\alpha(s)} = S_o(t)$$

como queríamos demostrar, entendiendose la última igualdad como notación.

La trivial verificación de las propiedades (1) a (4) de la proposición 1.2.1 nos lleva al siguiente resultado

Proposición 1.8.5. La función $S_o(t) = E\left[S(t)/r(t)\right]$ es una función de supervivencia no aleatoria.

De nuevo, al ser $S_o(t)$ una función de supervivencia no aleatoria, la consideraremos dentro del contexto Bayesiano en el que nos movemos, como la función de supervivencia "a priori", nombre que nos servirá para referirnos a ellos de ahora en adelante, llamando a la correspondiente función de distribución $F_o(t) = 1 - S_o(t)$, función de distribución "a priori".

Teorema 1.8.1. Sea $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$ un proceso gamma extendido. Entonces,

$$E\left[S^2(t)\right] = E\left[(S(t/r(t)))^2\right] = \exp\left[-\int_{[0,t)} \log\left[1+2\beta(s)\cdot(t-s)\right] d\alpha(s)\right]$$

en donde la primera igualdad se entiende como notación.

En efecto:

$$E\left[S^2(t)\right] = E\left[e^{-2\int_{[0,t)} r(s)ds}\right] = \lim_{n\to\infty} E\left[e^{-2\int_{[0,t)} r_n(s)ds}\right]$$

siendo $r_n(s)$ las variables aleatorias anteriormente definidas. Así pues, utilizando los cálculos efectuados en el apartado anterior,

$$E\left[S^2(t)\right] = \lim_{n\to\infty} \prod_{i=2}^{1'} E\left[e^{-2\beta(t_i)\cdot(t-t_i)\left[Z(t_i)-Z(t_{i-1})\right]}\right] =$$

$$= e^{-\int_{[0,t)} \log\left[1+2\beta(s)\cdot(t-s)\right]d\alpha(s)}$$

como queríamos demostrar.

<u>Teorema 1.8.2</u>. Sea $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$ un proceso gamma extendido. Entonces, si $x < y$,

$$E\left[S(x).S(y)\right] = E\left[S(y).S(x)\right] =$$

$$= e^{-\int_{[0,x)} \log\left[1+\beta(s)(x+y-2s)\right]d\alpha(s) - \int_{[x,y)} \log\left[1+\beta(s)(y-s)\right]d\alpha(s)}$$

En efecto:

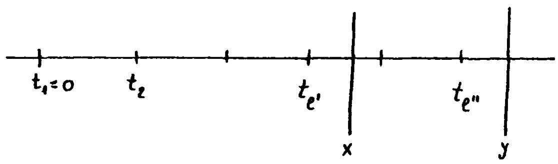
$$E\left[S(x).S(y)\right] = E\left[e^{-\left(\int_{[0,x)} r(t)dt + \int_{[0,y)} r(s)ds\right)}\right] =$$

$$= \lim_{n\to\infty} E\left[e^{-\int_{[0,x)} r_n(t)dt - \int_{[0,y)} r(s)ds}\right]$$

y si $t_{1'}$ es el mayor t_i menor que x y $t_{1''}$ el mayor menor que y, será

$$E\left[S(x).S(y)\right] = \lim_{n\to\infty} E\left[e^{-\sum_{i=2}^{1'} \beta(t_i)\left[Z(t_i)-Z(t_{i-1})\right].(x-t_i) - \sum_{i=2}^{1''} \beta(t_i)\cdot \left[Z(t_i)-Z(t_{i-1})\right](y-t_i)}\right]$$

Como $x < y$, la situación será de la siguiente forma



y por tanto,

$$E[S(x).S(y)] = \lim_{n\to\infty} E\left[e^{-\sum_{i=2}^{l'} \beta(t_i)(x+y-2t_i)[Z(t_i)-Z(t_{i-1})] - \sum_{i=l'+1}^{l''} \beta(t_i)(y-t_i)[Z(t_i)-Z(t_{i-1})]}\right] =$$

$$= (\lim_{n\to\infty} \prod_{i=2}^{l'} E\left[e^{-\beta(t_i)(x+y-2t_i)[Z(t_i)-Z(t_{i-1})]}\right])$$

$$(\lim_{n\to\infty} \prod_{n=l'+1}^{l''} E\left[e^{-\beta(t_i)(y-t_i)[Z(t_i)-Z(t_{i-1})]}\right]$$

$$= (\lim_{n\to\infty} \prod_{i=2}^{l'} (\frac{1}{1+\beta(t_i)(x+y-2t_i)})^{\Delta\alpha_i})(\lim_{n\to\infty} \prod_{i=l'+1}^{l''} (\frac{1}{1+\beta(t_i)(y-t_i)})^{\Delta\alpha_i})$$

$$= e^{-\int_{[0,x)} \log[1+\beta(s)(x+y-2s)]d\alpha(s) - \int_{[x,y)} \log[1+\beta(s)(y-s)]d\alpha(s)}$$

como queríamos demostrar.

<u>Corolario 1.8.1</u>. Sea $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$, un proceso gamma extendido. Entonces, dados $x,y \in R^+$ es

$$E[S(x).S(y)] = \begin{cases} e^{-\int_{[0,y)} \log[1+\beta(s)(x+y-2s)]\,d\alpha(s) - \int_{[y,x)} \log[1+\beta(s)(x-s)]\,d\alpha(s)} & , \text{ si } \; y < x \\ e^{-\int_{[0,x)} \log\,[1+2\beta(s)(x-s)]\,d\alpha(s)} & , \text{ si } \; y = x \\ e^{-\int_{[0,x)} \log[1+\beta(s)(x+y-2s)]\,d\alpha(s) - \int_{[x,y)} \log[1+\beta(s)(y-s)]\,d\alpha(x)} & , \text{ si } \; x < y \end{cases}$$

y $E[S(x)] = e^{-\int_{[0,x)} \log\,(1+\beta(s)(x-s))\,d\alpha(s)}$

Este resultado se deduce facilmente de los teorems 1.8.1 y 1.8.2 y de la proposición 1.8.4.

<u>Función de covarianza del proceso gamma extendido</u>:

Si $x,y \in R^+$ la función de covarianza $r(x,y)$ viene definida por

$$r(x,y) = E[(S(x)-S_0(x))(S(y)-S_0(y))] = E[S(x).S(y)] -$$

$$- E[S(x)] \,.\, E[S(y)]$$

con lo que a partir del corolario 1.8.1 obtenemos la siguiente proposición:

<u>Proposición 1.8.6</u>. La función de covarianza $r(x,y)$ de un proceso gamma extendido $S(t/r(t)) \equiv \Gamma(\alpha(.),\, \beta(.))$ es

$$
r(x,y) = \begin{cases}
\cdot\exp\left[-\int_{[0,y)} \log[1+\beta(s)(x+y-2s)]\,d\alpha(s) - \int_{[y,x)} \log[1+\beta(s)(x-s)]\,d\alpha(s)\right] - \\
\quad - \exp\left[-\int_{[0,x)} \log[1+\beta(s)(x-s)]\,d\alpha(s) - \int_{[0,y)} \log[1+\beta(s)(y-s)]\,d\alpha(s)\right], \\
\quad \text{si} \quad y < x \\
\cdot\exp\left[-\int_{[0,x)} \log[1+2\beta(s)(x-s)]\,d\alpha(s)\right] - \exp\left[-2\int_{[0,x)} \log[1+\beta(s)(x-s)]\,d\alpha(s)\right], \\
\quad \text{si} \quad y = x \\
\cdot\exp\left[-\int_{[0,x)} \log[1+\beta(s)(x+y-2s)]\,d\alpha(s) - \int_{[x,y)} \log[1+\beta(s)(y-s)]\,d\alpha(s)\right] - \\
\quad - \exp\left[-\int_{[0,x)} \log[1+\beta(s)(x-s)]\,d\alpha(s) - \int_{[0,y)} \log[1+\beta(s)(y-s)]\,d\alpha(s)\right], \\
\quad \text{si} \quad x < y
\end{cases}
$$

Continuidad en media cuadrática del proceso gamma exponencial:

La función de covarianza en la diagonal es

$$
r(x,x) = \exp\left[-\int_{[0,x)} \log[1+2\beta(s)(x-s)]\,d\alpha(s)\right] -
$$
$$
- \exp\left[-2\int_{[0,x)} \log[1+\beta(s)(x-s)]\,d\alpha(s)\right]
$$

y sabemos que una condición suficiente para que un proceso sea continuo en media cuadrática es que $r(x,x)$ sea continua, cosa que aquí es cierta. Así pues, hemos probado el siguiente resultado.

Proposición 1.8.7. Todo proceso gamma extendido $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$ es continuo en media cuadrática.

Obsérvese que al verificarse ésta proposición, todos los momentos de $S(t/r(t))$ existirán.

Integración en media cuadrática del proceso gamma extendido:

Sea $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$ un proceso gamma extendido. Sea $Y = \int_0^\infty S(t)dt$ la variable aleatoria la cual nos indica la integral de Lebesgue del proceso $S(t) = S(t/r(t))$. (Como se ve, a veces abreviamos $S(t/r(t))$ por $S(t)$ por razones de comodidad).

Al igual que hicimos con los procesos gamma exponenciales y con los procesos homogéneos simples daremos a continuación algunos teoremas que más tarde, cuando hagamos estimaciones, nos serán de mucha utilidad. Todos ellos se deducen facilmente de los teoremas 1.3.4, 1.3.5, 1.3.6 y 1.3.7, así como de las observaciones efectuadas en cada caso.

Teorema 1.8.3. Sea $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$ un proceso gamma extendido, con función de distribución a priori asociada $F_0(t)$. Sea $g(t) : [0,\infty) \longrightarrow R$ una función medible de la variable aleatoria T asociada a $S(t)$, tal que

$$\int_F (\int_0^\infty |g(t)|dF(t))dP(F) < \infty.$$

Entonces, $\int_F (\int_0^\infty g(t)dF(t))dP(F)$ es la esperanza de $g(T)$ respecto de la distribución $F_0(t)$.

En efecto:

Por verificarse el toerema de Fubini, será

$$\int_F \left(\int_0^\infty g(t)dF(t)\right)dP(F) = \int_0^\infty g(t) \left(\int_F dF(t)dP(F)\right) =$$

$$= \int_0^\infty g(t)dF_o(t) = E[g(T)/F_o]$$

ya que $S_o(t) = \int_{F'} S(t)\, dP(S) \Longrightarrow F_o(t) = \int_F F(t)dP(F) \Longrightarrow$

$\Longrightarrow dF_o(t) = \int_F dF(t)dP(F).$

Corolario 1.8.2. Sea $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$ un proceso gamma extendido, con función de distribución a priori asociada $F_o(t)$. Entonces $E[Y]$ es el momento de primer orden de $F_o(t)$.

El resultado es inmediato tomando $g(T) = T$ y aplicando el teorema anterior.

Sin más que aplicar el teorema 1.3.6 obtenemos el siguiente resultado.

Teorema 1.8.4. Sea $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$ un proceso gamma extendido con función de supervivencia a priori asociada $S_o(t)$. Sea $g(t) : [0,\infty) \longrightarrow R$ una función medible de T tal que

$$\int_{F'} \left(\int_0^\infty |S(t)g(t)|dt\right)dP(S) < \infty.$$

Entonces

$$\int_{F'} \left(\int_0^\infty S(t)g(t)dt\right)dP(S) = \int_0^\infty g(t)S_o(t)dt$$

Para terminar, damos tres resultados que son transcripción de los teoremas 1.3.4, 1.3.5 y 1.3.7 al ser el proceso gamma extendido continuo en media cuadrática. Se completan los resultados utiliel corolario 1.8.1.

Teorema 1.8.5. Sea $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$ un proceso gamma extendido. Sea $g : [0,\infty) \longrightarrow R^+ \cup \{0\}$ una función medible no negativa tal que $\forall n \in N$ es

$$\int_0^n g^2(s)ds < \infty.$$

Supongamos que

$$\int_{F'} (\int_0^\infty g(t)S(t)dt)(\int_0^\infty g(u)S(u)du)dP(S) < \infty.$$

Entonces, es

$$\int_{F'} (\int_0^\infty g(t)S(t)dt)(\int_0^\infty g(u)S(u)du)dP(S) =$$

$$= \int_0^\infty g(t)(\int_0^t \exp\left[-\int_{[0,u)} \log[1+\beta(s)(u+t-2s)]d\alpha(s) -\right.$$

$$\left. - \int_{[u,t)} \log\ [1+\beta(s)(t-s)]d\alpha(s)\right] . g(u)du)dt +$$

$$+ \int_0^\infty g(t)(\int_t^\infty g(u)\ \exp\ \left[-\int_{[0,t)} \log[1+\beta(s)(u+t-2s)]d\alpha(s) -\right.$$

$$\left. - \int_{[t,u)} \log[1+\beta(s)(u-s)]d\alpha(s)\right]du)dt$$

Teorema 1.8.6. Sea $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$ un proceso gamma extendido. Sean $g_1 : [0,\infty) \longrightarrow R^+ \cup \{0\}$ y $g_2 : [0,\infty) \longrightarrow R^+ \cup \{0\}$ dos funciones medibles no negativas tales que $\forall n \in N$ es

$$\int_0^n g_1(s)ds < \infty \qquad y \qquad \int_0^n g_2(s)ds < \infty.$$

Supongamos que

$$\int_{F'} \left(\int_0^\infty g_i(t)S(t)dt\right)\left(\int_0^\infty g_j(u)S(u)du\right)dP(S) < \infty \quad \text{para} \quad \begin{array}{l} i=1,2, \\ j=i,2. \end{array}$$

Entonces, es

$$\int_{F'} \left(\int_0^\infty g_1(t)S(t)dt\right)\left(\int_0^\infty g_2(u)S(u)du\right)dP(S) =$$

$$= \int_0^\infty g_1(t)\left(\int_0^t \exp\left[-\int_{[0,u)} \log\left[1+\beta(s)(u+t-2s)\right]d\alpha(s) -\right.\right.$$

$$\left.\left. - \int_{[u,t)} \log\left[1+\beta(s)(t-s)\right]d\alpha(s)\right] \cdot g_2(u)du\right)dt +$$

$$+ \int_0^\infty g_1(t)\left(\int_t^\infty \exp\left[-\int_{[0,t)} \log\left[1+\beta(s)(u+t-2s)\right]d\alpha(s) -\right.\right.$$

$$\left.\left. - \int_{[t,u)} \log\left[1+\beta(s)(u-s)\right]d\alpha(s)\right] \cdot g_2(u)du\right)dt$$

<u>Teorema 1.8.7</u>. Sea $S(t/r(t)) \equiv \Gamma(\alpha(.), \beta(.))$ un proceso gamma extendido. Sean $h_1 : [0,\infty) \longrightarrow \mathbb{R}$, y $h_2 : [0,\infty) \longrightarrow \mathbb{R}$ dos funciones de la variable aleatoria T, tales que sus derivadas con respecto a la medida de Lebesgue sean respectivamente $g_1(x) = \dfrac{dh_1(x)}{dx}$ y $g_2(y) = \dfrac{dh_2(y)}{dy}$, las cuales supondremos son funciones medibles no negativas y que verifican la condición

$$\forall n \in \mathbb{N}, \quad \int_0^n g_i(s)ds < \infty \quad , \quad i=1,2.$$

Supongamos que $c_1 = -\lim\limits_{x\to 0} h_1(x) < \infty$ y que $c_2 = -\lim\limits_{x\to 0} h_2(x) < \infty$.

Supongamos por último que

$$\int_{F'} \left(\int_0^\infty |g_i(t)S(t)|dt\right)dP(S) < \infty \quad , \quad i=1,2$$

y que

$$\int_{F'} \left(\int_0^\infty g_i(t)S(t)dt\right)\left(\int_0^\infty g_j(u)S(u)du\right)dP(S) < \infty, \quad i=1,2, \quad j=1,2.$$

Entonces,

$$\int_{F} \left(\int_0^\infty h_1(x)\,dF(x)\right)\left(\int_0^\infty h_2(y)\,dF(y)\right)dP(F) =$$

$$= c_1 \cdot c_2 - c_1 \int_0^\infty g_2(y) . \exp\left[-\int_{[0,y)} \log\ (1+\beta(s)(y-s))\,d\alpha(s)\right]dy\ -$$

$$- c_2 \int_0^\infty g_1(x)\ .\ \exp\left[-\int_{[0,x)} \log\ [1+\beta(s)(x-s)]\,d\alpha(s)\right]dx\ +$$

$$+ \int_0^\infty g_1(x)\left(\int_0^x \exp\left[-\int_{[0,y)} \log[1\ +\beta(s)(y+x-2s)]\,d\alpha(s)\ -\right.\right.$$

$$\left.\left. - \int_{[y,x)} \log[1+\beta(s)(x-s)]\,d\alpha(s)\right]\ .\ g_2(y)dy\right)dx$$

$$+ \int_0^\infty g_1(x)\left(\int_x^\infty \exp\left[-\int_{[0,x)} \log[1+\beta(s)(x+y-2s)]\,d\alpha(s)\ -\right.\right.$$

$$\left.\left. - \int_{[x,y)} \log[1+\beta(s)(y-s)]\,d\alpha(s)\right]\ g_2(y)dy\right)dx$$

CAPITULO 2

ESTIMACION CON DISTRIBUCION A PRIORI UN PROCESO GAMMA EXPONENCIAL

I. INTRODUCCION

Sea $T \geq 0$ una variable aleatoria real valorada con función de distribución desconocida F. Dada una muestra aleatoria simple $(T_1,\ldots,T_n)$ de T, lo que pretenderemos será el hacer estimaciones de una función $g(F)$, en donde g es una función definida sobre el espacio $\mathcal{F}$ de todas las distribuciones de probabilidad sobre $(\mathbb{R},\mathbb{B})$.

Supongamos que sobre el espacio medible $(\mathcal{F},\sigma_{\mathcal{F}})$, en donde $\sigma_{\mathcal{F}}$ es una σ-álgebra con respecto a la cual g sea medible, existe una medida de probabilidad $\mathcal{P}$, y consideremos la pérdida cuadrática

$$L(F,d) = \left[g(F) - d\right]^2$$

Si queremos hacer inferencias sobre $g(F)$, basándonos en $(T_1,\ldots,T_n)$, la regla de decisión Bayes será, Berger (1980) pag. 110,

$$d(x_1,\ldots,x_n) = \int_{\mathcal{F}} g(F)\, d\mathcal{P}_{x_1,\ldots,x_n}(F)$$

supuesta existente. Es decir, la media de la función $g(F)$ con respecto a la probabilidad a posteriori $\mathcal{P}_{T_1,\ldots,T_n}$ de F dada la muestra aleatoria $(T_1,\ldots,T_n)$.

Como hemos visto en el teorema 1.4.1, en el caso de que $\mathcal{P}$ sea la inducida por un proceso de Dirichlet en (R,B), la probabilidad a posteriori es manejable y tratable analíticamente. Desgraciadamente, parece ser éste el único caso en el que una probabilidad a priori en los problemas de decisión Bayesiana no paramétrica, resulta ser analíticamente manejable a posteriori. Proponemos aquí una posible via de solución del problema original de estimar la función $g(F)$ salvando la dificultad de ser la distribución a posteriori poco manejable: aproximar la solución del problema original de la búsqueda de la regla Bayes para $g(F)$, restringiendo el espacio de reglas de decisión al conjunto de combinaciones lineales de algún conjunto dado de funciones de la muestra. Consideraremos la función de pérdida cuadrática y estudiaremos situaciones en las que haya datos censurados y situaciones en las que no los haya.

En el presente capítulo consideraremos el caso en el que la medida de probabilidad a priori P, es la inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$, proposición 1.6.5. En ésta situación, F se podrá escribir, definición 1.6.2, como

$$F(t) = 1 - e^{-\Lambda_o(t)} \quad , \qquad t \geq 0$$

en donde $\Lambda_o(t)$ es una variable aleatoria gamma $G(c\Lambda^*(t),c)$, con lo que

$$E[\Lambda_o(t)] = \frac{c\ \Lambda^*(t)}{c} = \Lambda^*(t)$$

y

$$Var(\Lambda_o(t)) = \frac{c\ \Lambda^*(t)}{c^2} = \frac{\Lambda^*(t)}{c}$$

Así pues, $\Lambda^*(t)$ puede considerarse en el contexto Bayesiano, y así lo haremos, como nuestro conocimiento a priori sobre $\Lambda_o(t)$, siendo c (y por tanto la función $l(c)$) un parámetro que nos indica la confianza en nuestro conocimiento a priori. Así, cuando $c \to \infty$ $(l(c) \to 1)$ se tendrá que $Var(\Lambda_o(t)) \to 0$, lo cual nos indica que la variable aleatoria $\Lambda_o(t)$, para cada t fijo, está degenerada en el número real $\Lambda^*(t)$ y nuestro conocimiento a priori será máximo. Por otro lado, si $c \to 0$ $(l(c) \to 0)$ será $Var(\Lambda_o(t)) \to \infty$ y nuestro conocimiento a priori será mínimo. Obsérve sin embargo, que dar a priori funciones Λ^* resulta bastante complejo en cuanto a su falta de realismo; una persona puede considerar como distribución a priori la uniforme $U(0,1)$ por ejemplo, pero dificilmente se le ocurriría la correspondiente $\Lambda^*(t)$,

$$\Lambda^*(t) = \begin{cases} 0 & \text{si} \quad t < 0 \\ \dfrac{\log(1-t)}{c \log \frac{c}{c+1}} & \text{si} \quad 0 \leq t < 1 \\ +\infty & \text{si} \quad t \geq 1 \end{cases}$$

como función Λ^* a priori. Es decir, como sabemos por la proposición 1.6.7, existe una distribución a priori marginal de la muestra $F_o(t) = 1 - S_o(t)$ relacionada con $\Lambda^*(t)$ por la expresión, proposición 1.6.7,

$$F_o(t) = 1 - \left(\frac{c}{c+1}\right)^{c\Lambda^*(t)}$$

que nos permitirá introducir conocimientos a priori de un modo más eficaz, permaneciendo el parámetro c con la misma significación que antes: $c \to \infty$ $(l(c) \to 1)$, conocimiento a priori óptimo; $c \to 0$ $(l(c) \to 0)$, conocimiento a priori pésimo.

Obsérvese que $F_o(t)$ es el parámetro del proceso neutral por la derecha, ya que

$$E_P[F(t)] = 1 - E_P[S(t)] = 1 - S_o(t) = F_o(t)$$

y por lo tanto, caracteriza a dicho proceso.

Así pues, resumiendo ésta breve introducción, en éste capítulo nos propondremos el encontrar estimadores Bayes de $g(F)$, así como el riesgo Bayes de $\mathcal{P}$ asociado, bajo las siguientes hipótesis:

(a) $\mathcal{P}$ es la probabilidad inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$.

(b) La función de pérdida es cuadrática.

(c) Nos restringiremos en su búsqueda a la clase de las reglas de decisión que son combinaciones lineales de ciertas funciones de la muestra, las cuales especificaremos en cada caso.

Vamos ahora a explicar un hecho que se repetirá constantemente a lo largo de las estimaciones de este capítulo y de los dos restantes. Es el hecho de por que los coeficientes que estimemos suman uno.

<u>Teorema 2.1.1</u>. Sea $\theta \in \Theta$ un parámetro que se quiere estimar. Sea π una distribución de probabilidad a priori sobre Θ tal que $E_\pi[\theta] \neq 0$. Entonces, si $E_\pi[E_m[X_i]] = E_\pi[\theta]$, $i=1,\dots,n$, siendo E_m la esperanza con respecto a la distribución muestral de la variable aleatoria X_i, la regla Bayes para θ, bajo pérdida cuadrática, dentro de la clase de estimadores de la forma

$$a_1 X_1 + \dots + a_n X_n$$

es tal que las estimaciones $\hat{a}_1,\dots,\hat{a}_n$ verifican que

$$\hat{a}_1 + \hat{a}_2 + \dots + \hat{a}_n = 1$$

En efecto:

Habrá que determinar las $a_1,\dots,a_n$ que hagan mínimo el riesgo Bayes

$$\int_{\Theta} \int_{X} (\theta - a_1 X_1 - a_2 X_2 - \dots - a_n X_n)^2 \, dQ \, d\pi$$

Pero sabemos que si d es un valor central, el mínimo de

$$\int_{\Theta} \int_{X} (\theta - d)^2 \, dQ \, d\pi$$

se alcanza cuando $d = E_{\pi}[E_m[\theta]]$, con lo que los $\hat{a}_1, \dots, \hat{a}_n$ deberán verificar que

$$\hat{a}_1 X_1 + \dots + \hat{a}_n X_n = E_{\pi}[E_m[\theta]] = E_{\pi}[\theta]$$

y tomando esperanza en ambos miembros deberá ser

$$\hat{a}_1 E_{\pi}[E_m[X_1]] + \dots + \hat{a}_n E_{\pi}[E_m[X_n]] = E_{\pi}[\theta]$$

o lo que es lo mismo

$$\hat{a}_1 E_{\pi}[\theta] + \dots + \hat{a}_n E_{\pi}[\theta] = E_{\pi}[\theta]$$

de donde deberá ser

$$\hat{a}_1 + \dots + \hat{a}_n = 1$$

Una vez aclarado éste hecho, vamos a ver otro teorema que aclara èl porqué el estimador de la función de distribución $\hat{F}(t)$, es $1 - \hat{S}(t)$ en donde $\hat{S}(t)$ es el estimador de la función de supervivencia. ¿Será siempre la función de un estimador Bayes, el estimador Bayes de la misma función del parámetro?. La respuesta nos la da el siguiente teorema que pone limitaciones a dicha función. Volviendo a la estimación de la función de distribución que veremos en la sección VI, allí no emplearemos éste teorema y lo haremos directamente, aunque luego al final hacemos una reseña de éste hecho. Por último, aun que los teoremas que a continuación veremos pueden generalizarse cuando empleasemos como función base otro parámetro que no sea $S(t)$, dada la importancia de dicha función y el amplio estudio que a ella le hemos dedicado la tomaremos como "base de referencia".

Teorema 2.1.2. Si $\hat{S}(t)$ es la regla Bayes para $S(t)$, bajo pérdida cuadrática, buscada dicha regla dentro de la clase de reglas de decisión de la forma $a\ S_n(t) + b\ E[S(t)]$ y $g(x)$ es una función lineal en $\mathbb{R}$, entonces existe una clase de reglas de decisión que son combinación lineal de $E[S(t)]$ y $S_n(t)$ tal que $g(\hat{S}(t))$ es la regla Bayes, bajo la misma función de pérdida y misma distribución a priori, para $g(S(t))$, siendo el riesgo mínimo alcanzado, asociado a $g(\hat{S}(t))$, el mismo que el asociado a $\hat{S}(t)$, en donde dicha clase de reglas de decisión es de la forma

$$g\big(a\ S_n(t) + b\ E[S(t)]\big)$$

En efecto:

Sea $\hat{S}(t) = [1 - p_n(t)]\ S_n(t) - p_n(t)\ E[S(t)]$ la regla Bayes para $S(t)$ ya estudiada anteriormente en la sección III del capítulo anterior.

La regla Bayes para $g(S(t))$ dentro de la clase de reglas de decisión de la forma

$$g(a\ S_n(t) + b\ E[S(t)])$$

es la que haga mínimo el riesgo Bayes

$$\int_{F'}\int_0^{\infty} (g(S(t)) - g(a\ S_n(t) + b\ E[S(t)]))^2\ dQ(S_n(t))dP(S)$$

pero si $g(x)$ es una función lineal, será de la forma

$$g(x) = Ax + B$$

de donde el riesgo anterior será de la forma

$$A^2 \int_{F'} \int_0^\infty (S(t) - a\, S_n(t) - b\, E[S(t)])^2 \, dQ(S_n(t))dP(S)$$

y el a y b que lo hacen mínimo serán los ya conocidos

$$\hat{a} = 1 - p_n(t) \quad y \quad \hat{b} = p_n(t)$$

y por tanto evidentemente el riesgo mínimo alcanzado será el mismo.

¿Qué ocurriría si tuviesemos dos parámetros?, ¿se mantendría éste resultado?. La respuesta, dada por el siguiente teorema, es que si, aunque ahora la función ha de ser bilineal.

Teorema 2.1.3. Sea $\hat{S}(t)$ la regla Bayes para $S(t)$, bajo pérdida cuadrática, buscada dicha regla dentro de la clase de reglas de decisión de la forma

$$a\, S_n(t) + b\, E[S(t)].$$

Sea $\hat{S}'(t)$ la regla Bayes para $S'(t)$, bajo la misma función de pérdida, buscada dicha regla dentro de la clase de reglas de decisión de la forma

$$a'\, S'_m(t) + b'\, E[S'(t)].$$

Supondremos que tanto $S(t)$ y $S'(t)$ como las muestras de tamaños n y m extraidas para obtener respectivamente $S_n(t)$ y $S'_m(t)$ son independientes. Si $g(x,y)$ es una forma bilineal en R, entonces existe una clase de reglas de decisión que son combinación lineal de $S_n(t)$, $S'_m(t)$, $E[S(t)]$ y $E[S'(t)]$, tal que $g(\hat{S}(t), \hat{S}'(t))$ es la regla Bayes, bajo la misma pérdida cuadrática, para $g(S(t),S'(t))$, siendo dicha clase de reglas de decisión, la clase

$$g(a\ S_n(t) + b\ E[S(t)],\ a'\ S'_m(t) + b'\ E[S'(t)]).$$

En efecto:

El riesgo Bayes a hacer mínimo será

$$\iiiint (g(S(t),S'(t)) - g(a\ S_n(t) + b\ E[S(t)],\ a'\ S'_m(t) + b'\ E[S'(t)]))^2\ dQ\ dQ'\ dP(S)\ dP'(S')$$

Pero por ser $g(x,y)$ una forma bilineal, será de la forma (piénsese en el hiperplano de R^3, lo cual servirá en el caso de haber más de dos parámetros)

$$g(x,y) = Ax + By + c$$

con lo que el mencionado riesgo quedará de la forma

$$A^2 \iint (S(t) - a\ S_n(t) - b\ E[S(t)])^2\ dQ\ dP(S) +$$
$$+ B^2 \iint (S'(t) - a'\ S'_m(t) - b'\ E[S'(t)])^2\ .$$
$$.\ dQ'\ dP'(S) + 2\ AB\ (\iint (S(t) - a\ S_n(t) - b[E\ S(t)])dQ\ dP\ (S))$$
$$(\iint (S'(t) - a'\ S'_m(t) - b'\ E[S'(t)])\ dQ'\ dP'(S')).$$

Pero como $E_P[E_Q[S_n(t)]] = E_P[S(t)]$

y $\qquad E_P\Big[E_Q\big[E[S(t)]\big]\Big] = E_P[S(t)] = E[S(t)]$

podemos aplicar el teorema 2.1.1 y deberá de ser $a+b=1$, con lo que

$$\iint (S(t) - a\, S_n(t) - b\, E[S(t)])dQ\, dP = E[S(t)] - a\, E[S(t)] - b\, E[S(t)] = 0$$

por ser $a+b=1$, con lo que el doble producto desaparece quedando solo dos sumandos positivos que se harán mínimos cuando $\hat{a}$, $\hat{b}$, $\hat{a}'$, $\hat{b}'$ sean los correspondientes para obtener $\hat{S}(t)$ y $\hat{S}'(t)$, con lo que el resultado queda probado.

II. ESTIMACION DE LA FUNCION DE SUPERVIVENCIA CUANDO NO HAY DATOS CENSURADOS.

Consideremos el esquema expuesto en la sección anterior, cuando la probabilidad a priori P sobre el espacio paramétrico F, es la inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$ y con distribución de supervivencia a priori asociada $S_o(t)$.

En ésta sección calcularemos estimadores Bayes para $g(F)$ cuando fijado un $t \geq 0$, $g(F) = S(t) = 1 - F(t)$, es decir la función de supervivencia, desconocida, de la variable aleatoria T, siendo las estimaciones puntuales.

Supondremos como antes una función de pérdida cuadrática

$$L(S(t),d) = (S(t) - d)^2$$

y en la busqueda de la mencionada regla Bayes nos restringiremos a la clase de reglas de decisión que con combinación lineal de la función de supervivencia a priori $S_o(t)$ y de la función de supervivencia muestra $S_n(t)$. En concreto, el procedimiento que seguiremos se rá el siguiente: Tomaremos una muestra de tamaño n_1 de T y calcularemos la regla Bayes $d_1(t)$ dentro de la clase de reglas de decisión de la forma

$$a\, S_{n_1}(t) + b\, S_o(t)$$

A continuación tomaremos una muestra de tamaño n_2 de T y calcularemos la regla Bayes $d_2(t)$ dentro de la clase de reglas de decisión de la forma

$$a\, S_{n_2}(t) + b\, d_1(t)$$

y así sucesivamente, siendo a y b dos parámetros a determinar en cada caso. Así pues, buscaremos nuestra regla Bayes dentro de la clase de reglas de decisión que son combinación lineal de la información muestral contenida en $S_n(t)$ y de nuestro conocimiento a priori manifestado en $S_o(t)$, conocimiento éste que irá cambiando en el transcurso del experimento con la toma de nuevos datos y convirtiendose en $d_1(t)$, $d_2(t)$, etc. Estamos pues en un clásico análisis Bayesiano en donde nuestros conocimientos o creencias a priori $S_o(t)$ acerca del desconocido "verdadero estado de la Naturaleza" S(t), van siendo modificadas por los datos en el familiar análisis "priori-posteriori".

Estamos pues dentro del esquema expuesto en el teorema 1.3.9 y en sus preámbulos. Debemos hacer notar sin embargo que para ello deberán existir $E[S(t)]$ y $E[S^2(t)]$, cosa que podemos asegurar imponiendo la condición a $S_o(t)$ de ser una función de supervivencia no aleatoria, pero al ser P la inducida por un proceso gamma exponencial, ésta condición la tenemos asegurada, proposición 1.6.7 y podemos establecer por tanto el siguiente teorema como aplicación inmediata del teorema 1.3.9 y recordando que $E[S(t)] = S_o(t)$, proposición 1.6.6, y que $E[S^2(t)] = [S_o(t)]^{1(c)+1}$, teorema 1.6.2, estando ambas esperanzas calculadas con respecto a la probabilidad a priori P.

<u>Teorema 2.2.1</u>. Sea $T \geq 0$ una variable aleatoria observable la cual nos indica el tiempo de fallo de un sistema, con función de supervivencia desconocida $S(t)$. Si consideramos como probabilidad a priori P la inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t), c)$, y consecuentemente como distribución de supervivencia a priori $S_o(t) = (\frac{c}{c+1})^{c\Lambda^*(t)}$, la regla Bayes $\hat{S}(t)$, bajo pérdida cuadrática, para la función de supervivencia $S(t)$, siguiendo el procedimiento arriba expuesto, y siguiendo n el tamaño de la última muestra seleccionada es,

$$\hat{S}(t) = \frac{n[S_o(t)]^{1(c)} - n[S_o(t)]}{1+(n-1)[S_o(t)]^{1(c)} - n[S_o(t)]} S_n(t) +$$

$$+ \frac{1 - [S_o(t)]^{1(c)}}{1+(n-1)[S_o(t)]^{1(c)} - n[S_o(t)]} S_o(t)$$

y el riesgo Bayes asociado a P,

$$R_{min}(n) = \frac{-[S_o(t)]^2 + [S_o(t)]^{1(c)+2} + [S_o(t)]^{1(c)+1} - [S_o(t)]^{21(c)+1}}{1+(n-1)[S_o(t)]^{c} - n[S_o(t)]}$$

siendo $l(c) = \dfrac{\log \frac{c+2}{c+1}}{\log \frac{c+1}{c}}$.

Usualmente llamaremos

$$p_n(t) = \frac{1 - [S_o(t)]^{l(c)}}{1+(n-1)\,[S_o(t)]^{l(c)} - n\ S_o(t)}$$

y
$$1-p_n(t) = \frac{n\,[S_o(t)]^{l(c)} - n\ S_o(t)}{1+(n-1)\,[S_o(t)]^{l(c)} - n\ S_o(t)}$$

es decir a los coeficientes de $S_o(t)$ y $S_n(t)$ respectivamente. Obsérvese que dichos coeficientes suman uno, y no ha sido ésta una condición que le hayamos exigido, sino otra buena propiedad del estimador obtenido, que ahora podremos escribir más brevemente como

$$\hat{S}(t) = [1 - p_n(t)]\ S_n(t) + p_n(t)\ S_o(t)$$

Comportamiento asintótico muestral:

Vamos a ver a continuación que le ocurre a nuestro estimador cuando el tamaño de la muestra va hacia ∞. Para ello no tenemos más que aplicar el teorema 1.3.10; éste nos va a confirmar que $p_n(t)$ se va a ir haciendo cada vez más cercano a cero y $1-p_n(t)$ más cercano a uno, al ir aumentando el tamaño de la muestra. En otras palabras, cuando n va hacia ∞, la información a priori se pondera cada vez menos, es decir a la vez que n aumenta, van careciendo de importancia las diferencias entre diferentes estimaciones a priori $S_o(t)$ (a mayor tamaño muestral, menor importancia), quedando toda la información recogida en la muestra y por tanto en $S_n(t)$, el cual tenderá a $S(t)$

por ser un estimador consistente.

Teorema 2.2.2. En las mismas condiciones del teorema 2.2.1, sea $S(t)$ la función de supervivencia, desconocida, de la variable aleatoria T, y sea $\hat{S}(t)$ el estimador Bayes para $S(t)$, para una muestra aleatoria simple, determinado en el teorema anterior; es decir,

$$\hat{S}(t) = [1 - p_n(t)]\ S_n(t) + p_n(t)\ S_o(t).$$

Entonces, $\hat{S}(t)$ converge casi seguro hacia $S(t)$ cuando $n \to \infty$, para cada $t \geq 0$ fijo, es decir,

$$P\{|S(t) - S(t)| \longrightarrow 0, \quad \text{cuando} \quad n \to \infty\} = 1$$

Además, el correspondiente riesgo converge hacia cero cuando $n \to \infty$, también para cada $t \geq 0$.

Corolario 2.2.1. El estimador $\hat{S}(t) = [1 - p_n(t)]\ S_n(t) + p_n(t)\ S_o(t)$ converge en probabilidad y en ley hacia $S(t)$, y la función característica asociada a $\hat{S}(t)$ hacia la asociada a $S(t)$, cuando $n \to \infty$, para cada $t \geq 0$.

Corolario 2.2.2. El estimador $\hat{S}(t) = [1 - p_n(t)]\ S_n(t) + p_n(t)\ S_o(t)$ es un estimador consistente para $S(t)$, para una muestra aleatoria simple de tamaño n.

Comportamiento del estimador en función de c:

Ya hemos visto en la proposición 1.6.8 que el decir que c se acerca hacia cero o hacia ∞ es equivalente a decir que $1(c)$ se acerca hacia cero o hacia uno respectivamente, por lo que trabajaremos con $l(c)$ en vez de con c.

Como ya sabemos, para encontrar el estimador Bayes $\hat{S}(t)$ de $S(t)$, lo único que deberemos preguntar a la persona interesada en obtener una estimación de $S(t)$ es su estimación a priori $S_o(t)$ acerca de $S(t)$ y lo seguro o inseguro que él está acerca de ella, expresando dicho grado de confianza con un número $l(c)$, entre cero y uno: si él está muy seguro deberá dar a $l(c)$ un valor cercano a uno, si no está muy convencido de su estimación a priori $S_o(t)$ (cree que es bastante mala) deberá asignar, la persona interesada (un médico por ejemplo puede ser un importante usuario de ésta maquinaria inferencial) un número cercano a cero; si no tiene ninguna información acerca de la bondad de su $S_o(t)$ sería aconsejable hacer $l(c) = \frac{1}{2}$ ($l(c)$ no informativo), pudiendo tomar $l(c)$ todos los posibles valores entre cero y uno según la fé que el interrogante tenga en su estimación a priori.

Ya hemos visto que en el caso de considerar muestras suficientemente grandes (una buena pregunta sería: ¿cuán grande debe de ser una muestra grande?) la estimación a priori $S_o(t)$ que nuestro usuario nos diese no sería demasiado trascendente y el riesgo en cualquier caso convergería hacia cero.

Nos planteamos ahora otra interesante cuestión: ¿cómo afecta la posible elección de $l(c)$, independientemente del tamaño muestral a nuestro estimador y a nuestro riesgo asociado?. La respuesta nos la da el teorema 2.2.3; antes veamos una proposición que necesitaremos para la demostración del mencionado teorema.

<u>Proposición 2.2.1</u>. Sea $S_o(t) = (\frac{c}{c+1})^{c\Lambda^*(t)}$ y $l(c) = \dfrac{\log \frac{c+2}{c+1}}{\log \frac{c+1}{c}}$.

Entonces,

(a) $\lim\limits_{c\to o} S_o(t) = \lim\limits_{c\to o} [S_o(t)]^{l(c)} = 1$

(b) $\lim\limits_{c\to\infty} S_o(t) = \lim\limits_{c\to\infty} [S_o(t)]^{l(c)} = e^{-\Lambda^*(t)}$

En efecto:

(a) $\lim\limits_{c\to o} c\,\Lambda^*(t) \log c = \lim\limits_{c\to o} \Lambda^*(t) \frac{\log c}{1/c} = \Lambda^*(t) \lim\limits_{c\to o} \frac{1/c}{-1/c^2} = 0$

con lo que

$$\lim_{c\to o} c^{c\Lambda^*(t)} = e^{\lim\limits_{c\to o} c\Lambda^*(t) \log c} = 1$$

y por tanto,

$$\lim_{c\to o} S_o(t) = \lim_{c\to} \frac{c^{c\Lambda^*(t)}}{(c+1)^{c\Lambda^*(t)}} = \frac{1}{1} = 1$$

Por otro lado,

$$\lim_{c\to o} [S_o(t)]^{l(c)} = \lim_{c\to o} (\frac{c+1}{c+2})^{c\Lambda^*(t)} = 1$$

(b) Si $\Lambda^*(t) \neq 0$, $\lim\limits_{c\to\infty} [S_o(t)] = \lim\limits_{c\to\infty} (1 - \frac{\Lambda^*(t)}{-c\Lambda^*(t)})^{-c\Lambda^*(t)} =$

$= e^{-\Lambda^*(t)}$ y si $\Lambda^*(t) = 0 \Longrightarrow \lim\limits_{c\to\infty} [S_o(t)] = \lim\limits_{c\to\infty} 1 = 1 = e^{-\Lambda^*(t)}$.

Por otro lado, si $\Lambda^*(t) \neq 0$, $\lim\limits_{c\to\infty} [S_o(t)]^{l(c)} =$

$$= \frac{\lim\limits_{c\to\infty} (1 - \frac{2\Lambda^*(t)}{-c\Lambda^*(t)})^{-c\Lambda^*(t)}}{\lim\limits_{c\to\infty} (1 - \frac{\Lambda^*(t)}{-c\Lambda^*(t)})^{-c\Lambda^*(t)}} = e^{-\Lambda^*(t)},$$ resultado que se mantiene

si $\Lambda^*(t) = 0$.

<u>Teorema 2.2.3</u>. En las mismas condiciones del teorema 2.2.1, sea $S(t)$ la función de supetvivencia, desconocida, de la variable aleatoria $T \geq 0$, y sea

$$\hat{S}(t) = [1 - p_n(t)]\ S_n(t) + p_n(t)\ S_o(t)$$

el estimador Bayes para $S(t)$, y

$$R_{min}(n) = \frac{-[S_o(t)]^2 + [S_o(t)]^{1(c)+2} + [S_o(t)]^{1(c)+1} - [S_o(t)]^{21(c)+1}}{1 + (n-1)\ [S_o(t)]^{1(c)} - n\ [S_o(t)]}$$

su riesgo mínimo asociado, anteriormente determinados. Entonces,

(a) $\lim\limits_{c \to o} \hat{S}(t) = S_n(t)$ y $\lim\limits_{c \to o} R_{min}(n) = 0$

(b) $\lim\limits_{c \to \infty} \hat{S}(t) = e^{-\Lambda^*(t)}$ y $\lim\limits_{c \to \infty} R_{min}(n) = 0$

En efecto:

Piénsese que $[S_o(t)]^{1(c)} = (\frac{c+1}{c+2})^{c\Lambda^*(t)}$ y que $S_o(t) = (\frac{c}{c+1})^{c\Lambda^*(t)}$

(a)

$$\lim_{c \to o} p_n(t) = \lim_{c \to o} \frac{1 - [S_o(t)]^{1(c)}}{1+(n-1)\ [S_o(t)]^{1(c)} - n[S_o(t)]} =$$

$$= \lim_{c \to 0} \frac{-[S_o(t)]^{l(c)} \Lambda^*(t) \log\frac{c+1}{c+2} - c\Lambda^*(t)\left(\frac{c+1}{c+2}\right)^{c\Lambda^*(t)-1} \frac{1}{(c+1)^2}}{(n-1)\left([S_o(t)]^{l(c)}\Lambda^*(t)\log\frac{c+1}{c+2} + c\Lambda^*(t)\left(\frac{c+1}{c+2}\right)^{c\Lambda^*(t)-1}\frac{1}{(c+2)^2}\right) - n\left(S_o(t)\Lambda^*(t)\log\frac{c}{c+1} + c\Lambda^*(t)\left(\frac{c}{c+1}\right)^{c\Lambda^*(t)-1}\frac{1}{(c+1)^2}\right)} =$$

"

$= 0$

con lo que $\lim\limits_{c \to 0} (1 - p_n(t)) = 1$

y por tanto,

$$\lim_{c \to 0} \hat{S}(t) = \lim_{c \to 0} ([1-p_n(t)] S_n(t) + p_n(t)\ S_o(t)) = S_n(t)$$

Por otro lado,

$$R_{min}(n) = S_o(t) \cdot p_n(t) \cdot ([S_o(t)]^{1(c)} - S_o(t)) \Longrightarrow$$

$$\lim_{c \to 0} R_{min}(n) = (\lim_{c \to 0} S_o(t))(\lim_{c \to 0} p_n(t))(\lim_{c \to 0} ([S_o(t)]^{1(c)} -$$

$$- S_o(t))) = 1.0.0 = 0$$

(b)

$$\lim_{c \to \infty} p_n(t) = \lim_{c \to \infty} \frac{1 - [S_o(t)]^{1(c)}}{1+(n-1)[S_o(t)]^{1(c)} - n[S_o(t)]} =$$

$$= \frac{1 - e^{-\Lambda^*(t)}}{1+(n-1)e^{-\Lambda^*(t)} - n\ e^{-\Lambda^*(t)}} = 1$$

con lo que $\lim\limits_{c \to \infty} (1 - p_n(t)) = 1-1 = 0$

y por tanto,

$$\lim_{c \to \infty} \hat{S}(t) = \lim_{c \to \infty} ([1 - p_n(t)]\ S_n(t) + p_n(t)\ S_o(t)) = e^{-\Lambda^*(t)}$$

Por otro lado,

$$\lim_{c\to\infty} R_{min}(n) = (\lim_{c\to\infty} S_o(t))(\lim_{c\to\infty} p_n(t))(\lim_{c\to\infty} [S_o(t)]^{l(c)} -$$

$$- \lim_{c\to\infty} S_o(t)) = e^{-\Lambda^*(t)} \,.\, 1 \,.\, (e^{-\Lambda^*(t)} - e^{-\Lambda^*(t)}) = 0$$

y el teorema queda demostrado.

Como se ve, éste teorema nos dice que cuando nuestro conocimiento a priori es pésimo $(c \to 0 \iff l(c) \to 0)$ el estimador Bayes se transforma en la función de supervivencia empírica, el usual estimador no paramétrico, y toda la información viene recogida en la muestra. Por el contrario si nuestro conocimiento a priori es óptimo, $(c \to \infty \iff l(c) \to 1)$ la regla Bayes es $\hat{S}(t) = e^{-\Lambda^*(t)}$ que es lo mejor que puede ser, ya que cuando $c \to \infty$, $\Lambda_o(t)$ está degenerada en $\Lambda^*(t)$, con lo que el mejor estimador para $S(t) = e^{-\Lambda_o(t)}$ cuando estamos seguros (o mejor, casi seguros) acerca del valor de $\Lambda_o(t)$, es e elevado a menos ese valor, es decir, $e^{-\Lambda^*(t)}$.

Por otro lado, recuerdese que hemos eliminado al cero y al uno como posibles valores para $S_o(t)$, ya que cuando alguno de éstos se presentase, se tendría que $S(t)$ sería degenerada o lo que es lo mismo $c \to \infty$, es decir nuestro conocimiento a priori resultaría exacto y realmente estaríamos en una situación paramétrica, mientras que cuando $c \to 0$ estaremos en una clásica situación no paramétrica. Así pues, el hacer $c \to 0$ equivale a estudiar nuestro problema desde un punto de vista no paramétrico clásico, no Bayesiano, y el hacer $c \to \infty$ el estudiar realmente un problema paramétrico. Como se ve el tratamiento desde un punto de vista Bayesiano engloba al no Bayesiano $(c \to 0)$ y de ahí que resulte más interesante. "

Comparación de la regla obtenida con la obtenida por Ferguson:

Ferguson (1979) determina la media de $S(t)$ con respecto a la distribución a posteriori de la inducida por un proceso gamma exponencial, y aunque practicamente inutilizable dada su complejidad, (remitimos al lector a la comprobación de dicha afirmación leyendo dicho artículo) en el caso de muestras de tamaño uno y muestras no censuradas ésta es, con nuestra notación,

$$E[S(t)/X=x] = \begin{cases} [S_0(t)]^{l(c)} & \text{si} \quad t < x \\ S_0(t)[S_0(x)]^{l(c)-1} l(c) & \text{si} \quad t \geq x \end{cases}$$

y aunque ya vimos al final de la sección III del capítulo 1 dicha comparación y lo que representaba, vemos que nuestro estimador dado por el teorema 2.2.1 es

$$\hat{S}(t) = \begin{cases} [S_0(t)]^{l(c)} & \text{si} \quad t \leq x \\ S_0(t) \left(\dfrac{1-[S_0(t)]^{l(c)}}{1 - S_0(t)}\right) & \text{si} \quad t > x \end{cases}$$

que como vemos coincide para $t < x$ y es muy próximo para $t > x$.

Por otro lado, Ferguson no determina ningún riesgo Bayes asociado, y además no consigue dar una interpretación clara de los parámetros que en el proceso intervienen.

En el estimador obtenido por Ferguson (1979), el tamaño esperado a posteriori del salto en una observación, depende de donde ocurra la observación: el tamaño del salto en x depende, en el estimador por él obtenido, de x a través de $S_o(x)$. En el obtenido en éste trabajo tal cosa no ocurre, con lo que cabría pensar en algún valor de c como "tamaño muestral a priori".

Otra muy buena propiedad del estimador lineal $\hat{S}(t)$ aquí determinado, es que cuando $c \to 0$, $\hat{S}(t)$ tiende hacia el estimador de máxima verosimilitud, $S_n(t)$, de $S(t)$, no importando cual sea el valor escogido para $S_o(t)$, situación ésta que no se cumple para el estimador de Ferguson.

Aunque apenas interesante, si conviene hablar para terminar, del problema sin muestra.

Teorema 2.2.4. El estimador Bayes para $S(t)$ (acción Bayes), respecto a la pérdida cuadrática, en el problema sin muestra, es la función de supervivencia a priori $S_o(t)$, siendo el riesgo Bayes mínimo asociado

$$R_{\min}(0) = [S_o(t)]^{1(c)+1} - [S_o(t)]^2$$

En efecto:

La regla Bayes para $S(t)$ (acción Bayes al no existir experimentación), será la que haga mínimo el riesgo Bayes

$$\int_{F'} (S(t) - d)^2 \, dP(S)$$

siendo $E_P[S(t)] = S_o(t)$ la que lo hace mínimo. En cuanto al riesgo Bayes mínimo, será

$$R_{\min}(0) = V(S(t)) = E_P[S^2(t)] - (E_P[S(t)])^2 = [S_o(t)]^{1(c)+1} - [S_o(t)]^2$$

como queríamos demostrar.

Corolario 2.2.3. El extraer una muestra de tamaño $n > 0$ es siempre preferido a abordar el problema sin muestra, en cuanto a la estimación de $S(t)$, por disminuir los riesgos.

En efecto, tal afirmación se sigue de ser

$$R_{min}(n) = p_n(t)\ R_{min}(0)$$

y $p_n(t) \leq 1$, verificandose la igualdad solamente cuando $c \to \infty$, que es cuando nuestra estimación a priori es perfecta.

Tasa de convergencia del riesgo:

Teorema 2.2.5. La tasa de convergencia del riesgo asociado a la regla Bayes $\hat{S}(t)$ es

$$\frac{R_{min}(n)}{R_{min}(n-1)} = \frac{n-1}{n} + \frac{p_n(t)}{n} = 1 - \frac{1}{n}\ (1 - p_n(t))$$

En efecto:

Como $R_{min}(m) = p_m(t)\ R_{min}(0)$, será

$$R_{min}(n) = \frac{p_n(t)}{p_m(t)}\ R_{min}(m)$$

En particular, si $m = n-1$,

$$\frac{R_{min}(n)}{R_{min}(n-1)} = \frac{p_n(t)}{p_{n-1}(t)} = 1 - \frac{1}{n}\ (1 - p_n(t)) = \frac{n-1}{n} + \frac{p_n(t)}{n}$$

Corolario 2.2.4. Al irse acercando n a $+\infty$, el cociente entre los riesgos se acerca a uno.

Dicha afirmación es evidente del hecho de que

$$\lim_{n \to \infty} (1 - p_n(t)) = 1$$

Al conocer la tasa de convergencia, podremos determinar si aumentamos un elementos muestral o no antes de realizar la muestra, en función de si vamos a ganar mucho o no, ya que $p_n(t)$ depende solo de n y de $S_o(t)$.

III. ESTIMACION DE LA FUNCION DE SUPERVIVENCIA CUANDO HAY DATOS CENSURADOS

Sea como siempre $T \geq 0$ una variable aleatoria la cual nos indica el tiempo de fallo de un sistema o de un individuo de una población homogénea. Supondremos que dicha variable aleatoria tiene asociada una función de supervivencia $S(t)$ que supondremos desconocida. Nuestro objetivo será el de determinar la regla Bayes para $S(t)$, bajo pérdida cuadrática, y considerando como distribución a priori la inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$. Evidentemente, en el problema sin muestra no puede haber datos censurados, por lo que el estudio efectuado en la sección anterior sigue siendo válido. Supongamos ahora que decidimos extraer una muestra de tamaño n, y que en ésta muestra hay individuos que por causas extrañas al experimento han fallecido, es decir, existen individuos que llamaremos censurados, que no han muertos por la causa que estamos estudiando, o lámparas que no lucen por haberse fundido, sino por haberse caído por ejemplo. En éstas circunstancias estamos en el caso de estimar la función de supervivencia cuando hay datos censurados. Buscaremos la regla Bayes dentro de la clase de reglas de decisión de la forma

$$a_1\left[S_n(t)\right]^{\frac{n-\delta}{n}} + a_2\left[S_n(t)\right]^{\frac{\delta}{n}} + b\ E\left[S(t)\right]$$

en donde n es el tamaño muestral, $S_n(t)$ la función de supervivencia muestral y δ el número de individuos censurados en dicha muestra.

Supondremos a δ independiente de T y con esperanza en el muestreo conocida $E[\delta] = \delta_o$.

Evidentemente cuando no hay individuos censurados en la muestra la clase de estimadores que estamos considerando aquí coincidirá con la estudiada en la sección anterior.

Estamos pues en las condiciones establecidas en el teorema 1.3.12, y recordando que cuando P es la inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$, es

$$E\left[S^2(t)\right] = \left[S_o(t)\right]^{1(c)+1} = \left(\frac{c}{c+2}\right)^{c\Lambda^*(t)}$$

y

$$E\left[S(t)\right] = S_o(t) = \left(\frac{c}{c+1}\right)^{c\Lambda^*(t)}$$

siendo $1(c) = \dfrac{\log \frac{c+2}{c+1}}{\log \frac{c+1}{c}}$ y que $S_o(t) \neq 0$ y $S_o(t) \neq 1$ por las razones antes mencionadas de no hacer degenerada a la variable aleatoria $S(t)$, tenemos el siguiente teorema.

<u>Teorema 2.3.1</u>. Sea $T \geq 0$ una variable aleatoria observable la cual nos indica el tiempo de fallo de un sistema, con función de supervivencia desconocida $S(t)$. Consideremos como probabilidad a priori

P la inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$, y consecuentemente como distribución de supervivencia a priori $S_o(t) = (\frac{c}{c+1})^{c\Lambda^*(t)}$. Sea $T_1,\dots,T_n$ una muestra de tamaño n de T en la cual hay δ individuos censurados, con δ independiente de T y con $E[\delta] = \delta_o$ un número conocido.

Entonces, la regla Bayes para $S(t)$, bajo pérdida cuadrática, buscada dicha regla dentro de la clase de estimadores de la forma

$$a_1[S_n(t)]^{\frac{n-\delta}{n}} + a_2[S_n(t)]^{\frac{\delta}{n}} + b\ S_o(t)$$

en donde $S_n(t)$ es la función de supervivencia muestral, es

$$\hat{a}_1[S_n(t)]^{\frac{n-\delta}{n}} + \hat{a}_2[S_n(t)]^{\frac{\delta}{n}} + \hat{b}\ S_o(t)$$

siendo $(\hat{a}_1,\hat{a}_2,\hat{b})$ el vector dado por

$$K(\hat{a}_1,\ \hat{a}_2,\ \hat{b})' = \vec{L}$$

siendo

$$K = \begin{bmatrix} -1+\frac{2}{n}\delta_o+\left(2-\frac{2\delta_o}{n}\right)S_o(t) & S_o(t) & \frac{\delta_o}{n}S_o(t)+\left(1-\frac{\delta_o}{n}\right)[S_o(t)]^2 \\ S_o(t) & 1-\frac{2}{n}\delta_o+\frac{2}{n}\delta_o\ S_o(t) & \left(1-\frac{\delta_o}{n}\right)S_o(t)+\frac{\delta}{n}[S_o(t)]^2 \\ \frac{\delta_o}{n}S_o(t)+\left(1-\frac{\delta_o}{n}\right)[S_o(t)]^2 & \left(1-\frac{\delta_o}{n}\right)S_o(t)+\frac{\delta_o}{n}[S_o(t)]^2 & [S_o(t)]^2 \end{bmatrix},$$

y

$$\vec{L} = \left(\frac{\delta_o}{n} S_o(t) + \left(1 - \frac{\delta_o}{n}\right)[S_o(t)]^{1(c)+1},\ \left(1 - \frac{\delta_o}{n}\right) S_o(t) + \right.$$
$$\left. + \frac{\delta_o}{n} [S_o(t)]^{1(c)+1},\ [S_o(t)]^2\right)'$$

Así mismo, el riesgo Bayes mínimo alcanzado (el asociado a P), es

$$R_{min}(n) = [S_o(t)]^{1(c)+1} - \vec{L}'\ (\hat{a}_1,\ \hat{a}_2,\ \hat{b})'$$

IV. ESTIMACION DEL TIEMPO MEDIO DE SUPERVIVENCIA.

Sea $T \geq 0$ una variable aleatoria observable real valorada con función de distribución asociada, desconocida, $F(t)$, y sea $S(t)$ la correspondiente función de supervivencia. Consideraremos que estamos en el mismo esquema expuesto en la introducción de este capítulo, cuando P es la probabilidad a priori, inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t), c)$, siendo $S_o(t)$ la función de supervivencia a priori asociada y $F_o(t) = 1 - S_o(t)$ la de distribución a priori.

Definición 2.4.1. Llamaremos tiempo medio de supervivencia, supuesto existente, correspondiente a la variable aleatoria T a

$$\mu(F) = \int_0^\infty t\ dF(t) = -\int_0^\infty t\ d\ S(t) = -S(t).t\Big|_0^\infty + \int_0^\infty S(t)dt =$$
$$= \int_0^\infty S(t)dt = \mu(S)$$

en donde la última igualdad se entiende como notación y la primera como definición.

Nos proponemos aquí el encontrar estimadores Bayes para $\mu(F)$, cuando la función de pérdida es cuadrática,

$$L(\mu(F),d) = (\mu(F) - d)^2$$

restringiendo su búsqueda a la clase de las reglas de decisión que son combinaciones lineales de ciertas funciones de la muestra que especificaremos en cada caso.

Por otro lado, los momentos respecto al origen de $F_o(t)$ son

$$\mu_o^i = \int_0^\infty t^i \, d\, F_o(t) = -\int_0^\infty t^i \, d\, S_o(t) = \int_F \int_0^\infty t^i \, dF(t)dP(F) \quad (17)$$

Supondremos existen todos los momentos de $F_o(t)$, lo que por (17) existirán equivalentemente todos los de $F(t)$. (Obsérvese que una condición suficiente para ello es suponer $\Lambda^*(t)$ continua en t).

En particular, si hacemos $i=1$ en (17), tendremos que

$$\mu_o^1 = \int_0^\infty t \, d\, F_o(t) = -\int_0^\infty t \, d\, S_o(t) = \int_F \int_0^\infty t \, d\, F(t)dP(F)$$

que llamaremos de ahora en adelante, tiempo medio de supervivencia a priori.

Teorema 2.4.1. El estimador Bayes para $\mu(F)$ (acción Bayes), respecto a la pérdida cuadrática, en el problema sin muestra, es el tiempo medio de supervivencia a priori μ_o^1, siendo el riesgo Bayes mínimo asociado,

$$R_{min}(0) = \int_0^\infty S_o(x)\left(\int_0^x [S_o(y)]^{1(c)} \, dy\right)dx + \\ + \int_0^\infty [S_o(x)]^{1(c)}\left(\int_x^\infty S_o(y)dy\right)dx - (\mu_o^1)^2$$

En efecto:

La regla Bayes (acción Bayes al no existir experimentación) para $\mu(F)$, será la que haga mínimo el riesgo Bayes

$$\int_F (\mu(F) - d)^2 \, dP(F)$$

siendo $d_o = E_P[\mu(F)]$ la regla que lo hace mínimo, es decir

$$d_o = \int_F \int_0^\infty t \; d\, F(t) = \mu_o^1$$

En cuanto al riesgo mínimo asociado, éste será

$$R_{min}(0) = V(\mu(F)) = E[\mu^2(F)] - (E[\mu(F)])^2,$$

pero

$$E[\mu^2(F)] = \int_F \left(\int_0^\infty u \; dF(u)\right)\left(\int_0^\infty v \; dF(v)\right) dP(F)$$

y aplicando el teorema 1.6.7 a las funciones $h_1(u) = u$ y $h_2(v) = v$, serán

$$c_1 = \lim_{u\to\infty} u \, e^{-\Lambda_o(u)} - \lim_{u\to o} u = 0$$

y

$$c_2 = \lim_{v\to\infty} v \, e^{-\Lambda_o(v)} - \lim_{v\to o} v = 0$$

con lo que, lo que más tarde llamaremos K_{11}, es

$$E[\mu^2(F)] = \int_0^\infty S_o(x)\left(\int_0^x [S_o(y)]^{1(c)} \, dy\right)dx +$$

$$+ \int_0^\infty [S_o(x)]^{1(c)} \left(\int_x^\infty S_o(y)dy\right)dx$$

y por tanto,

$$R_{min}(0) = \int_0^\infty S_o(x)(\int_0^x [S_o(y)]^{1(c)} dy)dx +$$

$$+ \int_0^\infty [S_o(x)]^{1(c)}(\int_x^\infty S_o(y)dy)dx - (\mu_o^1)^2$$

Como antes dijimos, vamos a buscar la regla Bayes para $\mu(F)$, cuando se extrae alguna muestra, en diferentes clases de reglas de decisión que son combinaciones lineales de funciones de la muestra. Consideraremos tres de ellas, aunque se podrían considerar muchas más.

Estimación primera:

Sea $T_1,\dots,T_n$ una muestra aleatoria simple, y $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} T_i$ la media muestral. Buscaremos en ésta estimación primera la regla Bayes para $\mu(F)$ dentro de la clase de reglas de decisión que son combinación lineal de la media muestral $\bar{x}$ y del tiempo medio de supervivencia a priori μ_o^1,

$$a\ \bar{x} + b\ \mu_o^1$$

Teorema 2.4.2. Sea $T \geq 0$ una variable aleatoria observable la cual nos indica el tiempo de fallo de un sistema, con función de distribución desconocida $F(t)$. Si consideramos como probabilidad a priori P la inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$, y consecuentemente como distribución de supervivencia a priori $S_o(t) = (\frac{c}{c+1})^{c\Lambda^*(t)}$, la regla Bayes $\hat{\mu}(S)$, bajo pérdida cuadrática, para el tiempo medio de supervivencia $\mu(S)$, dentro de la clase de reglas de decisión de la forma

$$a\ \bar{x} + b\ \mu_o^1$$

en donde $\bar{x}$ es la media muestral, μ_o^1 el tiempo medio de supervivencia a priori, y siendo n el tamaño muestral es,

$$\hat{\mu}(S) = \frac{n\, K_{11} - n(\mu_o^1)^2}{\mu_o^2 + (n-1)K_{11} - n(\mu_o^1)^2}\, \bar{x} + \frac{\mu_o^2 - K_{11}}{\mu_o^2 + (n-1)K_{11} - n(\mu_o^1)^2}\, \mu_o^1$$

y el riesgo Bayes asociado a P,

$$R_{min}(n) = \frac{\left[K_{11} - (\mu_o^1)^2\right] \cdot \left[\mu_o^2 - K_{11}\right]}{\mu_o^2 + (n-1)K_{11} - n(\mu_o^1)^2}$$

en donde $K_{11} = \int_0^\infty S_o(x) \left(\int_0^x \left[S_o(y)\right]^{1(c)} dy\right) dx +$

$+ \int_0^\infty \left[S_o(x)\right]^{1(c)} \left(\int_x^\infty S_o(y) dy\right) dx$ y $1(c) = \dfrac{\log \frac{c+2}{c+1}}{\log \frac{c+1}{c}}$.

En efecto:

Habrá que determinar el a y el b que hacen mínimo el riesgo Bayes

$$R = \int_F \int_{[0,\infty)^n} (\mu(S) - a\bar{x} - b\mu_o^1)^2\, d\,F(\bar{x}) dP(F) \qquad (18)$$

en donde por $dF(\bar{x})$ queremos representar $dF(\bar{x}) = dF(x_1) \cdot \ldots \cdot dF(x_n)$.

(18) será igual a

$$= b^2(\mu_o^1)^2 + a^2 \iint (\bar{x})^2\, dF(\bar{x}) dP(F) + \iint (\mu(S))^2 dF(\bar{x}) dP(F) +$$

$$+ 2ba\, \mu_o^1 \iint \bar{x}\, dF(\bar{x}) dP(F) - 2b\, \mu_o^1 \iint \mu(S) dF(\bar{x})\, dP(F) -$$

$$- 2a \iint \bar{x}\, \mu(S) dF(\bar{x}) dP(F),$$

con lo que

$$\frac{\partial R}{\partial a} = 2a \iint (\bar{x})^2 \, dF(\bar{x})dP(F) + 2b\, \mu_o^1 \iint \bar{x}\, dF(\bar{x})dP(F) -$$

$$- 2 \iint \bar{x}\, \mu(S)dF(\bar{x})dP(F)$$

$$\frac{\partial R}{\partial b} = 2(\mu_o^1)^2\, b + 2a\, \mu_o^1 \iint \bar{x}\, dF(\bar{x})dP(F) -$$

$$- 2\, \mu_o^1 \iint \mu(S)dF(\bar{x})dp(F)$$

Tenemos que

$$\iint \mu(S)dF(\bar{x})dP(F) = \int_F \mu(S)dP(F) = \mu_o^1.$$

$$\iint \bar{x}\, dF(\bar{x})dP(F) = \frac{1}{n}\int_F \int_0^\infty \cdots \int_0^\infty (x_1 +\ldots+ x_n)dF(x_1) \ldots$$

$$\ldots dF(x_n)\, dP(F) =$$

aplicando el teorema de Fubini,

$$= \frac{1}{n}\sum_{i=1}^{n} \int_F \left(\int_0^\infty x\, dF(x)\right)dP(F) = \int_F \mu(S)dP(F) = \mu_o^1.$$

$$\iint \bar{x}\, \mu(S)dF(\bar{x})dP(F) = \frac{1}{n}\int_F \int_0^\infty \cdots \int_0^\infty (x_1 +\ldots+ x_n)\mu(S)dF(x_1) \ldots$$

$$\ldots dF(x_n)dP(F) = \int_F \left(\int_0^\infty x\, dF(x)\right)\mu(S)dP(F) =$$

$$= \int_F \left(\int_0^\infty x\, dF(x)\right)\left(\int_0^\infty y\, dF(y)\right)dP(F)$$

y utilizando los cálculos del teorema 2.4.1,

$$= \int_0^\infty S_o(x)\left(\int_0^x [S_o(y)]^{1(c)}\, dy\right)dx + \int_0^\infty [S_o(x)]^{1(c)} \left(\int_x^\infty S_o(y)dy\right)dx =$$

$$= K_{11}.$$ "

$$\iint (\bar{x})^2 \, dF(\bar{x}) dP(F) = \frac{1}{n^2} \int_F \int_0^\infty \cdots \int_0^\infty \sum_{i=1}^{n} (x_i)^2 \, dF(x_1) \ \ldots \ dF(x_n)$$

$$dP(F) + \frac{2}{n^2} \int_F \int_0^\infty \cdots \int_0^\infty \sum_{i<j} x_i x_j \, dF(x_1) \ \ldots \ dF(x_n) dP(F) =$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} \int_F \left(\int_0^\infty x^2 dF(x)\right) dP(F) +$$

$$+ \frac{1}{n^2} n(n-1) \int_F \left(\int_0^\infty x \, dF(x)\right)\left(\int_0^\infty y \, dF(y)\right) dP(F) = \frac{1}{n} \mu_o^2 + \frac{n-1}{n} K_{11}$$

con lo que el sistema

$$\left.\begin{aligned} \frac{\partial R}{\partial a} &= 0 \\ \frac{\partial R}{\partial b} &= 0 \end{aligned}\right\}$$

se transforma en

$$\left.\begin{aligned} & a\left(\frac{\mu_o^2}{n} + \frac{n-1}{n} K_{11}\right) + b(\mu_o^1)^2 - K_{11} = 0 \\ & b(\mu_o^1)^2 + a(\mu_o^1)^2 - (\mu_o^1)^2 = 0 \end{aligned}\right\}$$

de donde

$$\hat{a} = \frac{n\ (K_{11} - (\mu_o^1)^2)}{\mu_o^2 + (n-1)\ K_{11} - n\ (\mu_o^1)^2}$$

y

$$\hat{b} = \frac{\mu_o^2 - K_{11}}{\mu_o^2 + (n-1)\ K_{11} - n\ (\mu_o^1)^2}$$

En cuanto al riesgo mínimo, será

$$R_{min}(n) = \hat{b}^2(\mu_o^1)^2 + \hat{a}^2\left(\frac{\mu_o^2}{n} + \frac{n-1}{n} K_{11}\right) + K_{11} + 2\hat{b}\hat{a}(\mu_o^1)^2 - 2\hat{b}(\mu_o^1)^2 -$$

$$- 2 \hat{a} K_{11}$$

por ser

$$\iint (\mu(S))^2 dF(\bar{x}) dP(F) = \int_F \left(\int_0^\infty x \, dF(x)\right)\left(\int_0^y dF(y)\right) dF(F) = K_{11}$$

con lo que

$$R_{min}(n) = (\hat{a})^2 \left(\frac{\mu_o^2}{n} + \frac{n-1}{n} K_{11} - (\mu_o^1)^2\right) + \hat{a}(2(\mu_o^1)^2 - 2 K_{11}) - ((\mu_o^1)^2 - K_{11})$$

$$= - \frac{n (K_{11} - (\mu_o^1)^2)}{\mu_o^2 + (n-1)K_{11} - n(\mu_o^1)^2} + (K_{11} - (\mu_o^1)^2) =$$

$$= \frac{(K_{11} - (\mu_o^1)^2)(\mu_o^2 - K_{11})}{\mu_o^2 + (n-1)K_{11} - n(\mu_o^1)^2} \; .$$

Conviene observar que $\hat{a} + \hat{b} = 1$.

Corolario 2.4.1. El estimador $\hat{\mu}(S) = \hat{a}\,\bar{x} + \hat{b}\,\mu_o^1$ converge cuando n tiende hacia ∞, hacia $\mu(S)$ casi seguro. Es decir, el estimador $\hat{\mu}(S)$ es consistente para $\mu(S)$ dentro de la clase de distribuciones con primer momento finito.

El resultado se obtiene al ser

$$\lim_{n\to\infty} \hat{b} = 0 \Longrightarrow \lim_{n\to\infty} \hat{a} = \lim_{n\to\infty} (1-\hat{b}) = 1$$

Es decir, cuando $n \to \infty$ el conocimiento a priori no se pondera y la única información es la contenida en la muestra.

Corolario 2.4.2. El riesgo asociado a $\hat{\mu}(S)$ converge hacia cero cuando $n \to \infty$.

El resultado es evidente por depender $R_{min}(n)$ de n solamente en el denominador.

Estimación segunda:

Sea de nuevo $T_1,\ldots,T_n$ una muestra aleatoria simple, y supongamos que esta vez, bajo pérdida cuadrática, queremos determinar la regla Bayes para $\mu(F)$ dentro de la clase de reglas de decisión de la forma

$$a_1 T_1 + a_2 T_2 + \ldots + a_n T_n + b\, \mu_o^1 \qquad (19)$$

pues bien, Goldstein (1975) demuestra que la regla Bayes para $\mu(F)$ dentro de la clase de reglas de decisión de la forma (19) es la misma que la regla Bayes en la clase de estimadores de la forma

$$a\, \bar{x} + b\, \mu_o^1$$

con lo que no consideraremos nunca éste tipo de estimaciones segunda y nos trasladaremos a la estimación primera cuando reglas como (19) se presenten.

Estimación tercera:

Ahora sólo consideraremos muestras de tamaño uno, y aunque veremos este tipo de estimación con más detalle en la próxima sección y los resultados que aquí daremos podrían ser deducidos como casos particulares de los de la siguiente sección, haremos un breve recorrido por el método.

Como siempre T será la variable y x un valor de ella. En concreto, queremos determinar la regla Bayes para $\mu(F)$, bajo pérdi da cuadrática, dentro de la clase de reglas de decisión de la forma

$$\lambda_o + \lambda_1 x + \lambda_2 x^2 + \ldots + \lambda_n x^n = (1, x, \ldots, x^n)\vec{\lambda}_n \qquad (20)$$

Teorema 2.4.3. En las condiciones del teorema anterior, si buscamos la regla Bayes para $\mu(S)$, bajo pérdida cuadrática, dentro de la clase de estimadores de la forma (20), ésta será

$$\hat{\mu}_{P_o}(S) = (1, x, x^2, \ldots, x^n)\ \hat{\vec{\lambda}}_n$$

y el riesgo Bayes asociado a P,

$$R_{min}(n) = K_{11} - \vec{B}_n'\ \hat{\vec{\lambda}}_n$$

en donde

$$\hat{\vec{\lambda}}_n = (\hat{\lambda}_o, \hat{\lambda}_1, \ldots, \hat{\lambda}_n)' = D_n^{-1}\ \vec{B}_n$$

siendo

$$\vec{B}_n = (K_{10}, K_{11}, \ldots, K_{1n})',$$

$F_o(t)$ la función de distribución a priori la cual supondremos tiene más de n puntos de incremento,

$$D_n = \begin{pmatrix} 1 & \mu_o^1 & \ldots & \mu_o^n \\ \mu_o^1 & \mu_o^2 & \ldots & \mu_o^{n+1} \\ \vdots & \ldots & \ldots & \ldots \\ \mu_o^{n+1} & \mu_o^{n+2} & \ldots & \mu_o^{2n} \end{pmatrix}$$

si $i \geq 1$, $j \geq 1$

$$K_{ji} = K_{ij} = i.j \int_0^\infty x^{i-1}\, S_o(x) \left(\int_0^x y^{j-1} [S_o(y)]^{1(c)} dy\right) dx +$$

$$+ \ i.j \int_0^\infty x^{i-1} \left[S_o(x)\right]^{1(c)} \left(\int_x^\infty y^{j-1} \ S_o(y)dy\right)dx$$

y $K_{io} = K_{oi} = \mu_o^i$, $i=0,1,\dots,n$.

En efecto:

El riesgo Bayes, a hacer mínimo será:

$$R = \int_F \int_0^\infty (\mu(S) - \sum_{i=o}^n x^i \lambda_i)^2 \ dF(x)dP(F)$$

$$= K_{11} - 2 \ \vec{\lambda}_n' \ . \ \vec{B}_n + \vec{\lambda}_n' \ D_n \ \vec{\lambda}_n \qquad (21)$$

ya que si $i \geq 1$, $j \geq 1$

$$\int_F \left(\int_0^\infty x^i dF(x)\right)\left(\int_0^\infty y^j dF(y)\right)dP(F) =$$

$$= i.j \int_0^\infty x^{i-1} \ S_o(x) \ \left(\int_0^x y^{j-1} \left[S_o(y)\right]^{1(c)} dy\right)dx \ +$$

$$+ \ i \ j \int_0^\infty x^{i-1} \left[S_o(x)\right]^{1(c)} \left(\int_x^\infty y^{j-1} \ S_o(y)dy\right)dx = K_{ij}$$

por el teorema 1.6.7 aplicado a las funciones $h_1(x) = x^i$ y $h_2(y) = y^j$, y por otro lado,

$$\int_F \left(\int_0^\infty x^i \ dF(x)\right) \ dP(F) = \mu_o^i = K_{io} = K_{oi}$$

De donde derivando en (21) se obtiene que

$$D_n \ \vec{\lambda}_n = \vec{B}_n$$

expresión esta que llamaremos ecuaciones ecuaciones normales y que nos permitirán determinar el vector $\vec{\lambda}_n$ que hace mínimo el riesgo Bayes, y que será por tanto,

$$\hat{\vec{\lambda}}_n = D_n^{-1} \vec{B}_n$$

(Obsérvese que $|D_n| \neq 0$ por tener F_o más de n puntos de incremento).

Por tanto el riesgo Bayes mínimo alcanzado será

$$R_{min}(n) = K_{11} - \vec{B}_n' D_n^{-1} \vec{B}_n = K_{11} - \vec{B}_n' \hat{\vec{\lambda}}_n$$

Corolario 2.4.3. La relación entre $R_{min}(n-1)$ y $R_{min}(n)$ es

$$R_{min}(n) - R_{min}(n-1) = (\hat{\lambda}_n^*)^2 \frac{|D_n|}{|D_{n-1}|}$$

en donde $\hat{\lambda}_n^*$ es el coeficiente estimado de x^n, dentro de la clase de reglas de decisión que son polinomios de grado n.

V. ESTIMACION DEL MOMENTO DE ORDEN i

Sea $T \geq 0$ una variable aleatoria observable, real valorada, la cual nos indica el tiempo de fallo de un sistema, con función de distribución desconocida $F(t)$ y función de supervivencia asociada $S(t)$. Consideraremos el esquema que ya viene siendo habitual a lo largo de todo este capítulo por ser P la probabilidad a priori, inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$, siendo $S_o(t)$ la función de supervivencia a priori asociada y $F_o(t)$ la de distribución a priori, de la cual supondremos existen todos sus momentos.

Calcularemos ahora el estimador Bayes para

$$\mu_i(F) = \int_0^\infty t^i\, dF(t) = -\int_0^\infty t^i\, dS(t) = i \int_0^\infty t^{i-1}\, S(t)dt = \mu_i(S)$$

en donde la primera y la última igualdad se entienden como notación. Supondremos como siempre la función de pérdida cuadrática,

$$L(\mu_i(F),d) = (\mu_i(F) - d)^2$$

<u>Teorema 2.5.1</u>. El estimador Bayes para $\mu_i(F)$ (acción Bayes al no existir experimentación), respecto a la pérdida cuadrática, en el problema sin muestra, es μ_o^i, el momento de orden i respecto al origen de $F_o(t)$, la función de distribución a priori, siendo el riesgo Bayes mínimo asociado,

$$R_{min}(0) = (i)^2 \int_0^\infty x^{i-1} S_o(x)\left(\int_0^x y^{i-1} [S_o(y)]^{1(c)} dy\right)dx +$$
$$+ (i)^2 \int_0^\infty x^{i-1} [S_o(x)]^{1(c)} \left(\int_x^\infty y^{i-1} S_o(y)dy\right)dx - (\mu_o^i)^2.$$

En efecto:

La regla Bayes (acción Bayes) será la que haga mínimo el riesgo Bayes

$$R = \int_F (\mu_i(F) - d)^2 \, dP(F)$$

que claramente es

$$E_P[\mu_i(F)] = \int_F \int_0^\infty x^i \, dF(x) \, dP(F) = \mu_o^i$$

es decir, el momento de orden i respecto al origen.

En cuanto al riesgo mínimo, éste será

$$R_{min}(0) = \mathrm{Var}\,(\mu_i(F)) = E[(\mu_i(F))^2] - (E[\mu_i(F)])^2$$

y como

$$E[(\mu_i(F))^2] = \int_F (\int_0^\infty x^i\, dF(x))\ (\int_0^\infty y^i\, dF(x))dP(F)$$

podemos aplicar el teorema 1.6.7 a las funciones

$$h_1(x) = x^i$$

$$h_2(y) = y^i$$

quedando

$$E[(\mu_i(F))^2] = (i)^2 \int_0^\infty x^{i-1}\ S_o(x)\ (\int_0^x y^{i-1}\ [S_o(y)]^{1(c)}\ dy)dx +$$

$$+ (i)^2 \int_0^\infty x^{i-1}\ [S_o(x)]^{1(c)}\ (\int_x^\infty y^{i-1}\ S_o(y)dy)dx$$

que más adelante llamaremos K_{ii}.

Así pues,

$$R_{min}(0) = (i)^2 \int_0^\infty x^{i-1}\ S_o(x)\ (\int_0^x y^{i-1}\ [S_o(y)]^{1(c)}\ dy)dx +$$

$$+ (i)^2 \int_0^\infty x^{i-1}\ [S_o(x)]^{1(c)}\ (\int_x^\infty y^{i-1}\ S_o(y)dy)dx - (\mu_o^i)^2$$

como queríamos demostrar.

Ahora consideraremos dos tipos de estimaciones, generalizaciones de las ya estudiadas para la media en la sección anterior, la primera de ellas para una muestra de tamaño n y la segunda para una de tamaño uno.

Estimación primera:

Restringiendo la busqueda del mencionado estimador Bayes a la clase de reglas de decisión de la forma

$$b_o \, \mu_o^i + b_1 \, \bar{\mu}_n^1 + b_2 \, \bar{\mu}_n^2 + \ldots + b_i \, \bar{\mu}_n^i$$

en donde

$$\bar{\mu}_n^j = \frac{1}{n} \sum_{h=1}^{n} (x_h)^j, \qquad j=1,2,\ldots,i$$

es el momento muestral de orden j, y suponiendo las hipótesis antes mencionadas obtenemos el siguiente teorema.

Teorema 2.5.1. Sea $T \geq 0$ una variable aleatoria observable la cual nos indica el tiempo de fallo de un sistema, con función de distribución desconocida $F(t)$. Si consideramos como probabilidad a priori P la inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$, y consecuentemente como distribución de supervivencia a priori $S_o(t) = (\frac{c}{c+1})^{c\Lambda^*(t)}$, la regla Bayes $\hat{\mu}_i(F)$, bajo pérdida cuadrática, para el momento de orden i $\mu_i(F)$, dentro de la clase de reglas de decisión de la forma

$$b_o \, \mu_o^i + b_1 \, \bar{\mu}_n^1 + b_2 \, \bar{\mu}_n^2 + \ldots + b_i \, \bar{\mu}_n^i = (\mu_o^i, \bar{\mu}_n^1, \bar{\mu}_n^2, \ldots, \bar{\mu}_n^i) \, \vec{b}_i,$$

es

$$\hat{\mu}_i(F) = (\mu_o^i, \bar{\mu}_n^1, \bar{\mu}_n^2, \ldots, \bar{\mu}_n^i) \, \hat{\vec{b}}_i$$

siendo el riesgo Bayes asociado a P

$$R_{min}(n) = K_{ii} - \vec{c}_i^{\,\prime} \, \hat{\vec{b}}_i$$

en donde $\hat{\vec{b}}_i = (\hat{b}_o, \hat{b}_1, \ldots, \hat{b}_i)' = K_i^{-1} \, \vec{c}_i$

$$\vec{c}_i = ((\mu_o^i)^2, K_{i1}, K_{i2}, \ldots, K_{ii})'$$

como antes

$$K_{ji} = K_{ij} = i.j \int_0^\infty x^{i-1}\, S_o(x)\Big(\int_0^x y^{j-1}\, [S_o(y)]^{1(c)} dy\Big)dx \; +$$

$$+\; i.j \int_0^\infty x^{i-1}\, [S_o(x)]^{1(c)} \Big(\int_x^\infty y^{j-1}\, S_o(y)dy\Big)dx \qquad j=1,\dots,i$$

y siendo

$$K_i = \begin{bmatrix} (\mu_o^i)^2 & \mu_o^i\,\mu_o^1 & \dots & (\mu_o^i)^2 \\ \mu_o^i\,\mu_o^1 & \frac{1}{n}\mu_o^2 + \frac{n-1}{n}K_{11} & \dots & \frac{1}{n}\mu_o^{i+1} + \frac{n-1}{n}K_{1i} \\ \dots & \dots & \dots & \dots \\ (\mu_o^i)^2 & \frac{1}{n}\mu_o^{i+1} + \frac{n-1}{n}K_{i1} & \dots & \frac{1}{n}\mu_o^{2i} + \frac{n-1}{n}K_{ii} \end{bmatrix}$$

que supondremos es tal que $|K_i| \neq 0$.

En efecto:

Deberemos determinar el vector $\vec{b}_i = (b_o, b_1, \dots, b_i)'$ tal que haga mínimo el riesgo Bayes

$$R = \int_F \int_0^\infty \dots \int_0^\infty \Big(\mu_i(S) - \sum_{j=1}^{i} b_i\, \bar{\mu}_n^j - b_o\, \mu_o^i\Big)^2 dF(x_1)\dots dF(x_n)dP(F)$$

$$= \iint (\mu_i(S))^2 dF(\bar{x})dP(F) + \sum_{j=1}^{i} (b_j)^2 \iint (\bar{\mu}_n^j)^2\, dF(\bar{x})dP(F) \; +$$

$$+\; (b_o)^2\, (\mu_o^i)^2 + 2 \sum_{j<h}^{i} b_j b_h \iint \bar{\mu}_n^j\, \bar{\mu}_n^h\, dF(\bar{x})\, dP(F) \; +$$

$$+ 2\, b_o\, \mu_o^i \sum_{j=1}^{i} b_j \iint \bar{\mu}_n^j\, dF(\bar{x})dP(F) - 2 \sum_{j=1}^{i} b_j \iint \mu_i(S)\bar{\mu}_n^j\, dF(\bar{x})dP(F)$$

$$- 2\, b_o\, \mu_o^i \int_F \mu_i(S)dP(F)$$

Tenemos que

$$\iint (\mu_i(S))^2 dF(\bar{x})dP(F) = \int_F (\int_0^\infty x^i\, dF(x))(\int_0^\infty y^i\, dF(y))dP(F) = K_{ii}$$

por aplicación del teorema 1.6.7 a las funciones $h_1(x) = x^i$, $h_2(y) = y^i$.

$$\int_F \int_0^\infty \cdots \int_0^\infty \bar{\mu}_n^j\, dF(\bar{x})\, dP(F) = \mu_o^j$$

$$\iint \bar{\mu}_n^j\, \bar{\mu}_n^h\, dF(\bar{x})\, dP(F) = \frac{1}{n}\, \mu_o^{h+j} + \frac{n-1}{n}\, K_{hj}$$

de nuevo aplicando el teorema 1.6.7 en éste caso a las funciones $h_1(x) = x^h$, $h_2(y) = y^j$

$$\iint (\bar{\mu}_n^j)^2\, dF(\bar{x})\, dP(F) = \frac{1}{n}\, \mu_o^{2j} + \frac{n-1}{n}\, K_{jj}$$

por ser un caso particular del anterior

$$\iint \mu_i(S)\, \bar{\mu}_n^j\, dF(\bar{x})dP(F) = \int_F (\int_0^\infty x^i dF(x))(\int_0^\infty y^j dF(y))dP(F) = K_{ij}$$

y por último,

$$\iint \mu_i(S)dF(\bar{x})dP(F) = \int_F (\int_0^\infty x^i\, dF(x))dP(F) = \mu_o^i$$

con lo que sustituyendo éstos cálculos en el riesgo a hacer mínimo obtenemos,

$$R = K_{ii} + \vec{b}_i'\, K_i\, \vec{b}_i - 2\, \vec{b}_i'\, \vec{c}_i$$

con lo que derivando e igualando a cero dichas derivadas obtendríamos

$$K_i \vec{b}_i = \vec{c}_i \Longrightarrow \hat{\vec{b}}_i = K_i^{-1} \vec{c}_i$$

obteniendose la regla Bayes

$$\hat{\mu}_i(S) = (\mu_o^i, \bar{\mu}_n^1, \ldots, \bar{\mu}_n^i) K_i^{-1} \vec{c}_i$$

como queríamos demostrar.

En cuanto al riesgo mínimo asociado, éste será

$$R_{min}(n) = K_{ii} - \vec{c}_i' K_i^{-1} \vec{c}_i = K_{ii} - \vec{c}_i' \hat{\vec{b}}_i \tag{22}$$

con lo que el teorema queda demostrado.

<u>Corolario 2.5.1</u>. Cuando $n \to \infty$, los coeficientes $\hat{b}_o, \hat{b}_1, \ldots, \hat{b}_{i-1}$ convergen a 0, y $\hat{b}_i$ a 1, es decir el vector $\hat{\vec{b}}_i$ converge al vector (0,0,...,0,1). Así pues, cuando $n \to \infty$, la única información es la contenida en el momento muestral de orden i.

<u>Corolario 2.5.2</u>. Cuando $n \to \infty$, $R_{min}(n) \to 0$, como facilmente se deduce del corolario 2.5.1 y de (22).

<u>Estimación segunda</u>:

Vamos ahora a generalizar para el momento de orden i, $\mu_i(F)$, la estimación realizada en la sección anterior, cuando nos restringimos en la búsqueda de la regla Bayes para $\mu_i(F)$ a la clase de reglas de decisión de la forma

$$\beta_o + \beta_1 x + \ldots + \beta_n x^n = (1, x, \ldots, x^n) \vec{\beta}_n \tag{23}$$

en donde x es el valor de una muestra de tamaño uno de la variable aleatoria T.

Utilizaremos la misma notación que en la "estimación tercera" de la anterior sección, y tenemos el siguiente resultado, suponiendo de nuevo que $F_o(t)$, la función de distribución a priori tiene más de n puntos de incremento.

<u>Teorema 2.5.2</u>. En las hipótesis del teorema anterior, si buscamos la regla Bayes para $\mu_i(S)$, bajo pérdida cuadrática dentro de la clase de los estimadores de la forma (23), ésta será

$$\hat{\mu}_i(S) = (1,\ x,\ x^2,\dots,x^n)\hat{\vec{\beta}}_n$$

siendo el riesgo Bayes mínimo asociado,

$$R_{min}(n) = K_{ii} - \vec{A}_n'\ \hat{\vec{\beta}}_n$$

en donde

$$\hat{\vec{\beta}}_n = (\hat{\beta}_o,\ \hat{\beta}_1,\dots,\hat{\beta}_n)' = D_n^{-1}\ \vec{A}_n$$

$$\vec{A}_n = (K_{io},\ K_{i1},\dots,K_{in})'$$

$$D_n = \begin{pmatrix} 1 & \mu_o^1 & \cdots & \mu_o^n \\ \mu_o^1 & \mu_o^2 & \cdots & \mu_o^{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_o^{n+1} & \mu_o^{n+2} & \cdots & \mu_o^{2n} \end{pmatrix}$$

$$K_{ij} = i.j \int_o^\infty x^{i-1}\ S_o(x)\left(\int_o^x y^{j-1}\ [S_o(y)]^{1(c)} dy\right)dx +$$

$$+\ ij \int_o^\infty x^{i-1}\ [S_o(x)]^{1(c)}\ \left(\int_x^\infty y^{j-1}\ S_o(y)dy\right)dx$$

y $\quad K_{io} = \mu_o^i \quad , \qquad j=1,\dots,n.$

En efecto:

El vector $\hat{\vec{\beta}}_n$ debe de hacer mínimo el riesgo Bayes

$$R = \int_F \int_0^\infty (\mu_i(S) - \sum_{i=o}^{n} x^i \beta_i)^2 \, dF(x) dP(F) =$$

$$= K_{ii} - 2 \vec{\beta}_n' \cdot \vec{A}_n + \vec{\beta}_n' D_n \vec{\beta}_n$$

con lo que derivando y haciendo cero dicha derivada obtenemos que

$$\hat{\vec{\beta}}_n = D_n^{-1} \vec{A}_n$$

resultando por tanto como riesgo mínimo

$$R_{min}(n) = K_{ii} - \vec{A}_n' D_n^{-1} \vec{A}_n = K_{ii} - \vec{A}_n' \vec{\beta}_n$$

Corolario 2.5.3. La relación entre $R_{min}(n-1)$ y $R_{min}(n)$ es

$$R_{min}(n) - R_{min}(n-1) = (\hat{\beta}_n^*)^2 \frac{|D_n|}{|D_{n-1}|}$$

en donde $\hat{\beta}_n^*$ es el coeficiente estimado para x^n, dentro de la clase de reglas de decisión que son polinomios de grado n.

VI. ESTIMACION DE LA FUNCION DE DISTRIBUCION

Supongamos el esquema habitual de ser $T \geq 0$ una variable aleatoria la cual nos representa el tiempo de fallo de un sistema. Sea $F(t)$ la función de distribución de T, la cual supondremos desconocida. Si P, la distribución de probabilidad a priori, suponemos es la inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t),c)$ y la función de pérdida es la cuadrática

$$L(F(t),d) = F(t) - d)^2$$

lo que pretenderemos hacer en ésta sección es el encontrar el estimador Bayes para $g(F) = F(t)$.

Supongamos en primer lugar el problema sin muestra.

Teorema 2.6.1. En las condiciones arriba expuestas, el estimador Bayes para $F(t)$ (acción Bayes al no haber experimentación), en el problema sin muestra, es la función de distribución a priori $F_o(t) = 1 - S_o(t) = 1 - (\frac{c}{c+1})^{c\Lambda^*(t)}$, siendo el riesgo mínimo asociado

$$R_{min}(0) = [S_o(t)]^{1(c)+1} - [S_o(t)]^2$$

En efecto:

$F_o(t) = E[F(t)]$ es quien hace mínimo el riesgo Bayes

$$\int_F (F(t) - d)^2 \, dP(F)$$

con lo que la primera parte ya está demostrada.

Por otro lado,

$$R_{min}(0) = V(F(t)) = V(1-S(t)) = V(S(t)) = [S_o(t)]^{1(c)+1} - [S_o(t)]$$

Como se ve a partir del teorema 2.2.4, en el problema sin muestra, tenemos el mismo riesgo estimando la función de distribución que la función de supervivencia, cosa que era de esperar.

Consideremos ahora el proceso consistente en buscar la regla Bayes para $F(t)$, tomando una muestra de tamaño n_1, dentro de la clase de regla de decisión de la forma

$$a_t F_{n_1}(t) + b_t F_o(t)$$

en donde $F_n(t) = 1 - S_n(t)$ es la función de distribución muestral y $F_o(t)$ la a priori, y llamemos a dicha regla $d_1(t)$. A continuación extraigamos otra muestra de tamaño n_2 y busquemos ahora la regla Bayes para $F(t)$, siempre bajo pérdida cuadrática, dentro de la clase de reglas de decisión de la forma

$$a_t \, F_{n_2}(t) + b_t \, d_1(t).$$

Si seguimos el proceso un número finito de etapas, y n es el tamaño muestral en la última etapa, tenemos el siguiente teorema.

<u>Teorema 2.6.2</u>. Sea $T \geq 0$ una variable aleatoria observable, la cual nos indica el tiempo de fallo de un sistema, con función de distribución desconocida $F(t)$. Consideremos como probabilidad a priori P la inducida por un proceso gamma exponencial de parámetros $(\Lambda^*(t), c)$, y consecuentemente como función de supervivencia a priori $S_o(t) = (\frac{c}{c+1})^{c\Lambda^*(t)}$ y función de distribución a priori $F_o(t) = 1 - S_o(t)$. Si $F_n(t)$ es la función de distribución muestral, la regla Bayes $\hat{F}(t)$, bajo pérdida cuadrática, para $F(t)$, siguiendo el procedimiento arriba expuesto, y siendo n el tamaño muestral en la última etapa es,

$$F(t) = \frac{n[S_o(t)]^{1(c)} - n[S_o(t)]}{1+(n-1)[S_o(t)]^{1(c)} - n[S_o(t)]} \, F_n(t) + $$

$$+ \frac{1 - [S_o(t)]^{1(c)}}{1+(n-1)[S_o(t)]^{1(c)} - n[S_o(t)]} \, F_o(t)$$

siendo el riesgo Bayes asociado a P

$$R_{min}(n) = \frac{-[S_o(t)]^2 + [S_o(t)]^{1(c)+2} + [S_o(t)]^{1(c)+1} - [S_o(t)]^{21(c)+1}}{1 + (n-1)\ [S_o(t)]^{1(c)} - n[S_o(t)]}$$

en donde $1(c) = \dfrac{\log \frac{c+2}{c+1}}{\log \frac{c+1}{c}}$.

En efecto:

Lo único que tenemos que hacer es determinar el a_t y b_t que hagan mínimo el riesgo Bayes

$$\int_F \int_0^{\infty} (F(t) - a_t\ F_n(t) - b_t\ F_o(t))^2\ dQ(F_n(t))dP(F)$$

y siguiendo el mismo camino que el expuesto en la sección II de éste capítulo se obtiene el resultado buscado, sin más que pensar que

$$E[F_n(t)] = F(t)$$

$$E[F_n^2(t)] = \frac{F(t)}{n} + \frac{n-1}{n} F^2(t)$$

$$E_p[F(t)] = F_o(t) = 1 - S_o(t)$$

$$E_p[F^2(t)] = E_p[1-2S(t) + S^2(t)] = 1 - 2S_o(t)+[S_o(t)]^{1(c)+}$$

En realidad se podrían deducir dos teoremas semejantes al 1.3.8 y 1.3.9 para funciones de distribución aleatorias generales y no necesariamente cuando intervienen los procesos gamma exponenciales. En concreto, que el estimador para la función de distribución F(t), buscado dicho estimador dentro de las reglas de decisión de la forma

$$a_t\ F_n(t) + b_t\ E[F(t)] \qquad (24)$$

es

$$\hat{F}(t) = \frac{n(E[S^2(t)] - (E[S(t)])^2)}{E[S(t)] + (n-1)\, E[S^2(t)] - n(E[S(t)])^2}\, F_n(t) +$$

$$+ \frac{E[S(t)] - E[S^2(t)]}{E[S(t)] + (n-1)E[S^2(t)] - n(E[S(t)])^2}\, E[F(t)]$$

resultado que se habría obtenido aunque en vez de $E[F(t)]$ hubiéramos puesto otra constante en (24). En la terminología habitual sería dicho estimador,

$$\hat{F}(t) = [1 - p_n(t)]\, F_n(t) + p_n(t)\, E[F(t)]$$

siendo como siempre

$$p_n(t) = \frac{E[S(t)] - E[S^2(t)]}{E[S(t)] + (n-1)\, E[S^2(t)] - n(E[S(t)])^2}$$

con lo que los resultados de convergencia también se cumplen.

En cuanto al riesgo, dado que $\hat{a}_t + \hat{b}_t = 1$, el riesgo mínimo será

$$\int_F \int_0^\infty (F(t) - \hat{a}_t\, F_n(t) - \hat{b}_t\, E[F(t)])^2\, dQ(F_n(t))dP(F) =$$

$$= \int_{F'} \int_0^\infty (1-\hat{a}_t - \hat{b}_t + \hat{a}_t\, S_n(t) + \hat{b}_t\, E[S(t)] - S(t))^2 dQ(S_n(t))dP(S) =$$

$$= \int_{F'} \int_0^\infty (S(t) - \hat{a}_t\, S_n(t) - \hat{b}_t\, E[S(t)])^2\, dQ(S_n(t))dP(S) =$$

$$= \frac{-(E[S(t)])^3 + E[S^2(t)]\, (E[S(t)])^2 + E[S(t)]\, E[S^2(t)] - (E[S^2(t)])^2}{E[S(t)] + (n-1)\, E[S^2(t)] - n\, (E[S(t)])^2}$$

como deducimos del teorema 1.3.8. Es decir, tenemos el mismo riesgo al estimar una función de supervivencia que la de distribución asociada, cosa que era de esperar.

En cuanto al caso particular que nos ocupa, al ser $S_o(t) \neq 0$ y $S_o(t) \neq 1$ para evitar la degeneración de la variable aleatoria $S(t)$, y observando que

$$E[S(t)] = S_o(t)$$
$$E[S^2(t)] = [S_o(t)]^{l(c)+1}$$

obtenemos que el riesgo mínimo alcanzado será

$$R_{min}(n) = \frac{-[S_o(t)]^2 + [S_o(t)]^{l(c)+2} + [S_o(t)]^{l(c)+1} - [S_o(t)]^{2l(c)+1}}{1 + (n-1)\,[S_o(t)]^{l(c)} - n\,[S_o(t)]}$$

como queríamos demostrar.

Así pues, si

$$p_n(t) = \frac{1 - [S_o(t)]^{l(c)}}{1 + (n-1)\,[S_o(t)]^{l(c)} - n\,[S_o(t)]}$$

nuestro estimador Bayes, en el caso de considerar como P la inducida por un proceso gamma exponencial, será

$$\hat{F}(t) = [1 - p_n(t)]\, F_n(t) + p_n(t)\, F_o(t)$$

<u>Comportamiento asintótico muestral</u>:

Como ya sabemos a partir de la sección II de éste capítulo, es

$$\lim_{n\to\infty} p_n(t) = 0 \qquad \text{y} \qquad \lim_{n\to\infty} (1 - p_n(t)) = 1$$

con lo que tenemos el siguiente teorema.

<u>Teorema 2.6.3</u>. En las mismas condiciones del teorema 2.6.2, el estimador Bayes $\hat{F}(t)$ allí obtenido para $F(t)$, converge casi seguro hacia $F(t)$ cuando $n \to \infty$ para cada $t \geq 0$ fijo, es decir,

$$P\{|\hat{F}(t) - F(t)| \longrightarrow 0, \quad \text{cuando} \quad n \to \infty\} = 1$$

Además, el correspondiente riesgo converge hacia cero cuando $n \to \infty$, para cada $t \geq 0$.

El resultado se obtiene en seguida del teorema 1.3.10 por ser $|\hat{F}(t) - F(t)| = |\hat{S}(t) - S(t)|$.

<u>Corolario 2.6.1</u>. El estimador $\hat{F}(t) = [1-p_n(t)]F_n(t) + p_n(t)\ F_o(t)$ converge en probabilidad y en ley hacia $F(t)$, y la función caracte rística asociada a $\hat{F}(t)$ hacia la asociada a $F(t)$, cuando $n \to \infty$, para cada $t \geq 0$.

<u>Corolario 2.6.2</u>. El estimador $\hat{F}(t) = [1 - p_n(t)]\ F_n(t) + p_n(t)\ F_o(t)$ es un estimador consistente para $F(t)$, para una muestra aleatoria simple de tamaño n.

<u>Comportamiento del estimador en función de c</u>:

Obsérvese que claramente, como ya habrá observado el lector, es

$$\hat{F}(t) = 1 - \hat{S}(t), \qquad \forall t \geq 0$$

en donde $\hat{F}(t)$ es el estimador Bayes para la función de distribución $F(t)$, y $\hat{S}(t)$ el estimador Bayes para la correspondiente función de supervivencia $S(t) = 1 - F(t)$. A partir de ésto, se tiene el siguien te resultado sin más que observar el teorema 2.2.3 y recordar que el riesgo mínimo alcanzado en la estimación de $F(t)$ es el mismo que el alcanzado en la estimación de $S(t)$, obteniendo las mismas conclusiones que allí se conseguían.

Teorema 2.6.4. En las mismas condiciones del teorema 2.6.2, el estimador Bayes para $F(t)$ allí determinado

$$\hat{F}(t) = [1 - p_n(t)]\ F_n(t) + p_n(t)\ F_o(t)$$

y su riesgo mínimo asociado

$$R_{min}(n) = \frac{-[S_o(t)]^2 + [S_o(t)]^{1(c)+2} + [S_o(t)]^{1(c)+1} - [S_o(t)]^{21(c)+1}}{1 + (n-1)\ [S_o(t)]^{1(c)} - n\ [S_o(t)]}$$

verifican que

(a) $\lim\limits_{c \to o} \hat{F}(t) = F_n(t)$ y $\lim\limits_{c \to o} R_{min}(n) = 0$

(b) $\lim\limits_{c \to \infty} \hat{F}(t) = 1 - e^{-\Lambda^*(t)}$ y $\lim\limits_{c \to \infty} R_{min}(n) = 0$

Tasa de convergencia del riesgo:

Como el riesgo en la estimación de $S(t)$ y de $F(t)$ es el mismo, a la vista del teorema 2.2.5 y del corolario 2.2.4 se obtiene el siguiente resultado.

Teorema 2.6.5. La tasa de convergencia del riesgo asociado a la regla Bayes $\hat{F}(t)$, la cual tiende a uno al ir n hacia $+\infty$, es

$$\frac{R_{min}(n)}{R_{min}(n-1)} = \frac{n-1}{n} + \frac{p_n(t)}{n}$$

Por último cabe hacer notar que muchos de los resultados aquí obtenidos hubieran quedado muy sencillos aplicando los teorema 2.1.1 y 2.1.2; especialmente aplicando éste último a la función $g(x) = 1-x$, evidentemente lineal.

VII. PROBLEMA DE LAS DOS MUESTRAS

Sean $F(t)$ y $G(t)$ dos funciones de distribución sobre $[0,\infty)$, y sean $X_1,\ldots,X_m$ una muestra extraida mediante F e $Y_1,\ldots,Y_n$ una muestra extraida mediante G. Consideremos el problema de estimar la probabilidad

$$\Delta = P\,\{X \leq Y\} = \int_0^\infty F(y)dG(y)$$

siendo X e Y las variables aleatorias con funciones de distribución F y G respectivamente, y en donde la primera igualdad se entiende como notación.

Supongamos que $F(x)$ es una función de distribución aleatoria, un proceso gamma exponencial de parámetros $\left(\Lambda^*(x),c\right)$ en concreto, y que $G(y)$ también es una función de distribución aleatoria, en éste caso un proceso gamma exponencial de parámetros $\left(\Lambda'(y),c'\right)$. Sea P la probabilidad inducida por el proceso de parámetros $\left(\Lambda^*(c),c\right)$ y P' la inducida por el de parámetros $\left(\Lambda'(y),c'\right)$. Supondremos que P y P' (y consecuentemente F y G) son independientes.

Llamaremos

$$\Delta_0 = \int_0^\infty F_0(y)dG_0(y)$$

en donde $F_0(t) = 1 - S_0(t)$ es la función de distribución a priori " para $F(t)$ y $G_o(t) = 1 - S_o'(t)$ la correspondiente a priori para $G(t)$, es decir,

$$F_o(t) = E_P[F(t)] = 1 - S_o(t) = 1 - \left(\frac{c}{c+1}\right)^{c\Lambda^*(t)}$$

$$G_o(t) = E_{P'}[G(t)] = 1 - S_o'(t) = 1 - \left(\frac{c'}{c'+1}\right)^{c'\Lambda'(t)}$$

<u>Teorema 2.7.1.</u> En las condiciones antes expuestas y bajo pérdida cuadrática, la regla Bayes para $\Delta = \int_0^\infty F(t)dG(t)$ es

$$\hat{\Delta}_{m,n} = \int \hat{F}(t)d\hat{G}(t) = p_m(t)\; p_n'(t)\Delta_o + p_m(t)(1-p_n'(t))\;.$$

$$.\;\frac{1}{n}\sum_{j=1}^{n} F_o(Y_j) + (1-p_m(t))p_n'(t)\;\frac{1}{m}\sum_{i=1}^{m}\left[1-G_o(X_i^-)\right] +$$

$$+\;(1-p_m(t))(1-p_n'(t))\;\frac{U}{n.m}$$

buscada dicha regla dentro del conjunto de reglas de decisión de la forma

$$a_1\,\Delta_o + a_2\,\frac{1}{n}\sum_{j=1}^{n} F_o(Y_j) + a_3\,\frac{1}{m}\sum_{i=1}^{m}\left[1-G_o(X_i^-)\right] + a_4\,\frac{U}{n.m}$$

en donde $p_m(t)$ y $p_n'(t)$ son especificadas más abajo, y siendo U = n° de pares (X_i,Y_j) para los cuales $X_i \leq Y_j$, es decir, el conocido estadístico utilizado por Mann y Whitney (1947).

En efecto:

Por el teorema 2.2.1 sabemos que las reglas Bayes para $S(t) = 1 - F(t)$ y $S'(t) = 1 - G(t)$ son, respectivamente

$$\hat{S}(t) = [1-p_m(t)]\; S_m(t) + p_m(t)\; S_o(t) = 1 - \hat{F}(t)$$

y $\quad \hat{S}'(t) = [1-p_n'(t)]\; S_n'(t) + p_n'(t)\; S_o'(t) = 1 - \hat{G}(t)$

siendo

$$p_m(t) = \frac{1 - [S_o(t)]^{1(c)}}{1+(m-1)[S_o(t)]^{1(c)} - m\, S_o(t)} \;;$$

$$S_m(t) = 1 - F_m(t) = \frac{n^\circ \text{ de } X_i > t}{m}$$

y
$$p_n'(t) = \frac{1 - [S_o'(t)]^{1(c')}}{1+(n-1)[S_o'(t)]^{1(c')} - n\, S_o'(t)};$$

$$S_n' = 1 - G_n(t) = \frac{n^\circ \text{ de } Y_i > t}{n}$$

Por otro lado, claramente, la función $g(S,S') =$ $= \int_o^\infty (1 - S(t))d(1 - S'(t))$ es bilineal, por lo que la regla Bayes para Δ, dentro de una clase de reglas de decisión que luego veremos, es por el teorema 2.1.3,

$$\hat{\Delta}_{m,n} = \int_o^\infty (1-\hat{S}(t))d(1-\hat{S}'(t)) = \int_o^\infty \hat{F}(t)d\hat{G}(t) =$$

$$= p_m(t)\; p_n'(t) \int_o^\infty F_o(t)dG_o(t) + p_m(t)(1-p_n'(t)) \int_o^\infty F_o(t)dG_n(t) +$$

$$+ (1-p_m(t))p_n'(t) \int_o^\infty F_m(t)dG_o(t) + (1-p_m(t))(1-p_n'(t)) \int_o^\infty F_m(t)dG_n(t) =$$

$$= p_m(t)\; p_n'(t)\Delta_o + p_m(t)(1-p_n'(t))\; \frac{1}{n} \sum_{j=1}^{n} F_o(Y_j) +$$

$$+ (1-p_m(t))p_n'(t)\, \frac{1}{m} \sum_{i=1}^{m} [1-G_o(X_i^-)] + (1-p_m(t))(1-p_n'(t))\, \frac{U}{n.m}$$

en donde U es el número de pares (X_i, Y_j) para los cuales

$X_i \leq Y_j$, por ser

$$\int_0^\infty F_o(t)dG_n(t) = \frac{1}{n} \sum_{j=1}^{n} F_o(Y_j)$$

$$\int_0^\infty F_m(t)dG_o(t) = \frac{1}{m} \sum_{i=1}^{m} [1 - G_o(X_i^-)]$$

$$\int_0^\infty F_m(t)dG_n(t) = \frac{1}{n} \sum_{j=1}^{n} F_m(Y_j) = \frac{1}{n.m} \sum_{j=1}^{n} \sum_{i=1}^{m} I_{(-\infty, Y_j]}(x_i) =$$

$$= \frac{U}{n.m}$$

Obsérvese que $\hat{\Delta}_{m,n}$ es una mixtura de cuatro estimadores de Δ, siendo sus coeficientes tales que

$$p_m(t)\ p_n'(t) + p_m(t)\ (1-p_n'(t)) + (1-p_m(t))p_n'(t) +$$
$$+ (1-p_m(t))\ (1-p_n'(t)) = 1$$

pudiendo considerarse a Δ_o como una estimación a priori de Δ.

En cuanto a la clase de reglas de decisión con respecto a la cual $\hat{\Delta}_{m,n}$ es la regla Bayes, ésta es

$$g(a\ S_m(t) + b\ S_o(t)),\quad a'\ S_m'(t) + b'\ S_o'(t)) =$$

$$= \int_0^\infty (1-a\ S_m(t) - b\ S_o(t))\ d(1 - a'\ S_n'(t) - b'\ S_o'(t)) =$$

$$= b\ b'\ \Delta_o + b\ a'\ \frac{1}{n} \sum_{j=1}^{n} F_o(Y_j) + a\ b'\ \frac{1}{m} \sum_{i=1}^{m} [1-G_o(X_i^-)] +$$

$$+ a.a'\ \frac{U}{n.m}$$

y como a, b, a' y b' son constantes, en general la clase de reglas de decisión será

$$a_1\,\Delta_o + a_2\,\frac{1}{n}\sum_{j=1}^{n} F_o(Y_j) + a_3\,\frac{1}{m}\sum_{i=1}^{m}\left[1 - G_o(X_i^-)\right] + a_4\,\frac{U}{n.m}$$

Comportamiento asintótico muestral:

Recuerdese que $\lim\limits_{m\to\infty} p_m(t) = 0$ y que $\lim\limits_{n\to\infty} p_n'(t) = 0$, con lo que al acercarse m y n a $+\infty$, es decir, al engrandarse las muestras de forma desmesurada, $\hat{\Delta}_{m,n}$ se convierte en $\frac{U}{n\ m}$ el usual estimador de Mann-Whitney, desapareciendo toda la información a priori recogida en $G_o(t)$ y $F_o(t)$.

Comportamiento en función de c y c':

Al estudiar el teorema 2.2.3 vimos los comportamientos de los coeficientes en función de c que ahora quedarán,

$\lim\limits_{c'\to 0} p_n'(t) = 0$	$\lim\limits_{c'\to 0} [1-p_n'(t)] = 1$	$\lim\limits_{c'\to\infty} p_n'(t) = 1$	$\lim\limits_{c'\to\infty} [1-p_n'(t)] = 0$
$\lim\limits_{c\to 0} p_m(t) = 0$	$\lim\limits_{c\to 0} [1-p_m(t)] = 1$	$\lim\limits_{c\to\infty} p_m(t) = 1$	$\lim\limits_{c\to\infty} [1-p_m(t)] = 0$

con lo que viendo la expresión adoptada por el estimador $\hat{\Delta}_{m,n}$, dada en el teorema 2.7.1, podrán combinarse diferentes casos posibles y sacar diferentes conclusiones, pensando siempre que un c cercano a cero indica una muy escasa confianza en el conocimiento a priori $F_o(t)$ (análogamente $c' \to 0$, escasa confianza en $G_o(t)$), y que por el contrario un $c \to \infty$ indica una fé muy elevada en nuestro conocimiento $F_o(t)$ $(c' \to \infty$, buena estimación a priori $G_o(t))$, como comentábamos en el teorema 2.6.4, siempre recordando por si hace falta que

$$\lim_{c\to o} F_o(t) = 0 \qquad \lim_{c\to\infty} F_o(t) = 1 - e^{-\Lambda^*(t)}$$

$$\lim_{c'\to o} G_o(t) = 0 \qquad \lim_{c'\to\infty} G_o(t) = 1 - e^{-\Lambda'(t)}$$

En particular, quisiera resaltar la situación en la que nuestros conocimientos a priori $F_o(t)$ y $G_o(t)$ no nos merecen apenas confianza ($c \to o$ y $c' \to o$). En ésta situación

$$\hat{\Delta}_{m,n} \xrightarrow[\substack{c \to o \\ c' \to o}]{} \frac{U}{n.m}$$

es decir nuestro estimador converge al usual estimador no paramétrico (no Bayesiano).

CAPITULO 3

ESTIMACION CON DISTRIBUCION A PRIORI UN PROCESO HOMOGENEO SIMPLE

I. INTRODUCCION

Supongamos que $T \geq 0$ es una variable aleatoria real valorada, la cual nos indica el tiempo de fallo de un sistema. Supondremos que la función de distribución de T, $F(t)$ es desconocida. El esquema de éste capítulo será bastante similar al anterior, estimaremos una función medible $g(F)$, definida sobre F, utilizando una función de pérdida cuadrática

$$L(g(F),\ d) = (g(F) - d)^2$$

y restringiremos la búsqueda de dicha estimación a un conjunto de reglas de decisión que sean combinación lineal de ciertas funciones de la muestra que especificaremos en cada caso. Hay sin embargo una notable diferencia: la distribución de probabilidad a priori P definida sobre F va a ser ahora la inducida por un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$, proposición 1.7.2.

De nuevo, la distribución a posteriori siendo la a priori un proceso homogéneo simple es inmanejable, y es por ésto por lo que aproximaremos las reglas Bayes, buscando la solución dentro de algún conjunto de funciones de la muestra.

Como hemos dicho, en éste capítulo consideraremos la situación en la cual P es la inducida por un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$ (recuérdese que por tanto ha de ser $c > 0$), es decir la función de distribución $F(t)$, desconocida, es decir, aleatoria, es un proceso homogéneo simple, con lo que la asociada función de supervivencia $S(t) = 1 - F(t)$ también lo será, y por tanto, por la definición 1.7.1, podremos escribirla como

$$S(t) = e^{-Y_t}, \qquad t \geq 0$$

en donde la función generatriz de momentos de Y_t, será

$$M_{Y_t}(\theta) = E\left[e^{-\theta Y_t}\right] = e^{c\Lambda^*(t) \int_0^\infty (e^{-\theta z} - 1)\, dN(z)}$$

siendo $N(.)$ la medida de Lévy asociada allí expresada. Como sabemos por la proposición 1.7.3 y teorema 1.7.1, es

$$E\left[S(t)\right] = e^{-\Lambda^*(t)} = S_o(t)$$

y $$V\left[S(t)\right] = E\left[S^2(t)\right] - (E\left[S(t)\right])^2 = \left[e^{-\Lambda^*(t)}\right]^{\frac{2c+1}{c+1}} - e^{-2\Lambda^*(t)}$$

con lo que cuando $c \to \infty$ $(m(c) \to 1)$, $V\left[S(t)\right] \longrightarrow 0$ y cuando $c \to 0$ $(m(c) \to 0)$, $V\left[S(t)\right] \longrightarrow e^{-\Lambda^*(t)} - e^{-2\Lambda^*(t)}$ que es el valor máximo que toma $V\left[S(t)\right]$ como función de c.

Así pues, consideraremos a $S_o(t) = e^{-\Lambda^*(t)}$, dentro del contexto Bayesiano en el que nos movemos, como nuestro conocimiento a priori acerca de $S(t)$ y a c, o equivalentemente a la función $m(c)$, como a un indicador de la fuerza en la creencia de dicho conocimiento a priori, ya que cuando $c \to \infty$ $(m(c) \to 1)$, la variable aleatoria $S(t)$ es degenerada en la función de supervivencia no aleatoria $S_o(t) = e^{-\Lambda^*(t)}$, y por tanto valores de c grandes (o consecuentemente, valores de $m(c)$ cercanos a uno) deberán de ser tomados cuando creamos que dicho conocimiento a priori es excelente. Por el contrario, como cuando $c \to 0$ la $V\left[S(t)\right]$ se hace máxima, valores cercanos a 0 para c (y equivalentemente para $m(c)$) deberán ser los que acompañen a la correspondiente $S_o(t)$, cuando dicha estimación a priori resulte poco fiable.

Por último, indicar que como allí se dice, los teoremas 2.1.2 y 2.1.3 son válidos para cualquier función de supervivencia aleatoria siempre que cumpla las condiciones ya mencionadas en lo referente a los momentos, y en particular serán válidos al considerar los procesos homogéneos simples.

Por tanto, el resumen de propósitos para éste capítulo puede ser el siguiente: el encontrar estimadores Bayes para $g(F)$, así como el riesgo Bayes asociado a $\mathcal{P}$, bajo las siguientes hipótesis:

(a) $\mathcal{P}$ es la probabilidad a priori, que supondremos inducida por un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$.

(b) La función de pérdida es cuadrática.

(c) Nos restringiremos en la busqueda de dicha regla Bayes a la clase de reglas de decisión que son combinación lineal de ciertas funciones de la muestra que especificaremos en cada caso.

II. ESTIMACION DE LA FUNCION DE SUPERVIVENCIA CUANDO NO HAY DATOS CENSURADOS.

Con el esquema expuesto en la sección anterior, supondremos ahora que la función $g(F)$ la cual se pretende estimar es, para cada t fijo,

$$g(F) = S(t) = 1 - F(t)$$

Supongamos primero que nos encontramos en el problema sin muestra.

<u>Teorema 3.2.1</u>. El estimador Bayes para $S(t)$ (acción Bayes), respecto a la función de pérdida cuadrática, en el problema sin muestra, es la función de supervivencia a priori $S_o(t) = e^{-\Lambda^*(t)}$, siendo el riesgo Bayes mínimo asociado

$$R_{min}(0) = [S_o(t)]^{m(c)+1} - [S_o(t)]^2 = e^{-\Lambda^*(t).\frac{2c+1}{c+1}} - e^{-2\Lambda^*(t)}$$

siendo $m(c) = \frac{c}{c+1}$ la función definida en 1.7.2.

En efecto:

La acción Bayes, bajo pérdida cuadrática, en el problema sin muestra será

$$E_P[S(t)] = S_o(t) = e^{-\Lambda^*(t)}$$

proposición 1.7.3, siendo el riesgo Bayes asociado a P,

$$R_{min}(0) = V(S(t)) = [S_o(t)]^{m(c)+1} - [S_o(t)]^2 =$$

$$= e^{-\frac{2c+1}{c+1}\Lambda^*(t)} - e^{-2\Lambda^*(t)}$$

como facilmente se deduce a partir del teorema 1.7.1.

Supongamos ahora que para estimar $S(t)$ decidimos realizar experimentación. Para ello seguimos el siguiente proceso: Extraemos una muestra de tamaño n_1 y buscamos la regla Bayes para $S(t)$ dentro de la clase de reglas de decisión de la forma

$$a\, S_{n_1}(t) + b\, S_o(t)$$

siendo $S_{n_1}(t)$ la función de supervivencia muestral, $S_o(t)$ la función de supervivencia a priori y $\hat{d}_1$ la regla Bayes así obtenida. A continuación extraemos otra muestra, ahora de tamaño n_2, y buscamos la regla Bayes para $S(t)$ dentro de la clase de reglas de decisión de la forma

$$a\, S_{n_2}(t) + b\, d_1$$

Si repetimos el proceso un número finito de etapas y n es el tamaño muestral en la última de ellas, la clase de reglas de decisión en la cual buscaríamos al final sería

$$a\, S_n(t) + b\, d_k$$

siendo d_k la regla Bayes en la etapa inmediatamente anterior. Si suponemos por último que tanto $E[S(t)]$ como $E[S^2(t)]$ existen y que $S_o(t) \neq 1$ y $S_o(t) \neq 0$ para evitar el que la variable $S(t)$ sea degenerada, tenemos el siguiente resultado.

Teorema 3.2.2. Sea $T \geq 0$ una variable aleatoria observable, la cual supondremos nos indica el tiempo de fallo de un sistema, con función de supervivencia $S(t)$ desconocida. Si consideramos como probabilidad a priori P la inducida por un proceso homogéneo simple de parámetros $(\Lambda^*(t), c)$, y consecuentemente como distribución a priori $S_o(t) = e^{-\Lambda^*(t)}$, la regla Bayes $S(t)$ para la función de supervivencia $S(t)$, bajo pérdida cuadrática y siguiendo el procedimiento arriba expuesto es,

$$S(t) = \frac{n[S_o(t)]^{m(c)} - n[S_o(t)]}{1+(n-1)[S_o(t)]^{m(c)} - n[S_o(t)]}\, S_n(t) +$$

$$+ \frac{1 - [S_o(t)]^{m(c)}}{1+(n-1)[S_o(t)]^{m(c)} - n[S_o(t)]}\, S_o(t)$$

y el riesgo Bayes asociado a $\mathcal{P}$,

$$R_{min}(n) = \frac{-[S_o(t)]^2 + [S_o(t)]^{m(c)+2} + [S_o(t)]^{m(c)+1} - [S_o(t)]^{2m(c)+1}}{1 + (n-1)\,[S_o(t)]^{m(c)} - n\,[S_o(t)]}$$

siendo $m(c) = \frac{c}{c+1}$.

En efecto:

La existencia de $E[S(t)]$ y $E[S^2(t)]$ la tenemos asegurada al ser $S_o(t)$ una función de supervivencia no aleatoria, con lo que podremos aplicar el teorema 1.3.9, siendo

$$E[S(t)] = S_o(t) = e^{-\Lambda*(t)}$$

$$\text{y} \qquad E[S^2(t)] = [S_o(t)]^{m(c)+1}$$

y el resultado quedará probado.

Si llamamos

$$q_n(t) = \frac{1 - [S_o(t)]^{m(c)}}{1+(n-1)[S_o(t)]^{m(c)} - n\,S_o(t)}$$

$$\text{y} \qquad 1-q_n(t) = \frac{n[S_o(t)]^{m(c)} - n\,[S_o(t)]}{1+(n-1)\,[S_o(t)]^{m(c)} - n[S_o(t)]}$$

con lo que el estimador acabado de calcular podrá escribirse como

$$\hat{S}(t) = [1 - q_n(t)]\; S_n(t) + q_n(t)\; S_o(t)$$

Obsérvese que la suma de los coeficientes de $S_n(t)$ y $S_o(t)$ es uno como tenía que ser a partir del teorema 2.1.1.

Comportamiento asintótico muestral:

Vamos a ver a continuación un teorema que nos va a decir como al irse incrementando el tamaño muestral n, va careciendo de importancia la información a priori contenida en $S_o(t)$ y solamente tiene valor la recogida en la muestra, es decir en $S_n(t)$, ya que

$$\lim_{n \to \infty} q_n(t) = 0$$

y

$$\lim_{n \to \infty} [1 - q_n(t)] = 1$$

El mencionado teorema que ahora vemos se obtiene como aplicación directa del teorema 1.3.10.

Teorema 3.2.3. En las mismas condiciones del teorema 3.2.2, el estimador Bayes

$$\hat{S}(t) = [1 - q_n(t)]\ S_n(t) + q_n(t)\ S_o(t)$$

converge casi seguro hacia $S(t)$ cuando $n \to \infty$, para cada $t \geq 0$ fijo, es decir

$$P\{|\hat{S}(t) - S(t)| \longrightarrow 0, \quad \text{cuando} \quad n \to \infty\} = 1$$

Además, para cada $t \geq 0$ fijo, el correspondiente riesgo converge hacia cero cuando $n \to \infty$.

Corolario 3.2.1. El estimador $\hat{S}(t) = [1 - q_n(t)]\ S_n(t) + q_n(t)\ S_o(t)$ converge en probabilidad y en ley hacia $S(t)$, y la función característica asociada a $\hat{S}(t)$ hacia la asociada a $S(t)$, cuando $n \to \infty$, para cada $t \geq 0$.

Corolario 3.2.2. El estimador $\hat{S}(t) = [1-q_n(t)] S_n(t) + q_n(t)\ S_o(t)$ es un estimador consistente para $S(t)$, para una muestra aleatoria simple de tamaño n.

Comportamiento del estimador $\hat{S}(t)$ en función de c:

Tratamos ahora de ver lo que le pasa a nuestro estimador cuando c se acerca a los valores extremos, es decir cuando $c \to 0$ (nuestro conocimiento a priori $S_o(t)$ es pésimo) y cuando $c \to \infty$ (nuestro conocimiento a priori $S_o(t)$ es óptimo). Dado que $c \to 0 \iff m(c) \to 0$ y que $c \to \infty \iff m(c) \to 1$, trabajaremos indistintamente con c o con $m(c)$, aunque el individuo interesado en éste estudio posiblemente expresará mejor su grado de confianza en $S_o(t)$ como un número entre cero y uno, es decir, mediante $m(c)$.

Dado que $S_o(t) = e^{-\Lambda^*(t)}$ y que $m(c) = \frac{c}{c+1}$ es claro que

$$\lim_{c \to o} S_o(t) = e^{-\Lambda^*(t)} \qquad y \qquad \lim_{c \to \infty} S_o(t) = e^{-\Lambda^*(t)}$$

$$\lim_{c \to o} [S_o(t)]^{m(c)} = \lim_{c \quad o} \left[e^{-\Lambda^*(t)}\right]^{\frac{c}{c+1}} = 1 \qquad y$$

$$\lim_{c \to \infty} [S_o(t)]^{m(c)} = e^{-\Lambda^*(t)}$$

con lo que tenemos el siguiente resultado:

Teorema 3.2.4. En las mismas condiciones del teorema 3.2.2, si

$$\hat{S}(t) = [1 - q_n(t)]\ S_n(t) + q_n(t)\ S_o(t)$$

$$y \qquad R_{min}(n) = \frac{-[S_o(t)]^2 + [S_o(t)]^{m(c)+2} + [S_o(t)]^{m(c)+1} - [S_o(t)]^{2m(c)+1}}{1 + (n-1)[S_o(t)]^{m(c)} - n\,[S_o(t)]}$$

son el estimador Bayes y riesgo mínimo allí determinados, entonces,

(a) $\lim_{c\to o} \hat{S}(t) = S_n(t) \qquad y \qquad \lim_{c\to o} R_{min}(n) = 0$

(b) $\lim_{c\to\infty} \hat{S}(t) = e^{-\Lambda^*(t)} = S_o(t) \qquad y \qquad \lim_{c\to\infty} R_{min}(n) = 0$

En efecto:

Como sabemos,

$$q_n(t) = \frac{1 - [S_o(t)]^{m(c)}}{1+(n-1)\,[S_o(t)]^{m(c)} - n\,S_o(t)}$$

con lo que

$$\lim_{c\to o} q_n(t) = 0 \qquad y \qquad \lim_{c\to\infty} q_n(t) = 1$$

y por tanto

$$\lim_{c\to o} [1-q_n(t)] = 1 \qquad y \qquad \lim_{c\to\infty} [1-q_n(t)] = 0$$

de donde

$$\lim_{c\to o} \hat{S}(t) = \lim_{c\to o} ([1-q_n(t)]\, S_n(t) + q_n(t)\, S_o(t)) = S_n(t)$$

$$y \qquad \lim_{c\to\infty} \hat{S}(t) = \lim_{c\to\infty} ([1-q_n(t)]\, S_n(t) + q_n(t)\, S_o(t)) = e^{-\Lambda^*(t)}$$

Por otro lado,

$$R_{min}(n) = q_n(t)\ .\ ([S_o(t)]^{m(c)} - [S_o(t)]).S_o(t)$$

con lo que

$$\lim_{c\to o} R_{min}(n) = 0 \qquad y \qquad \lim_{c\to\infty} R_{min}(n) = 0$$

y el resultado queda probado.

Como se ve a la luz del teorema acabado de demostrar, cuando nuestro conocimiento a priori es pésimo $(c \to o)$ nuestro estimador "anula" la parte apriorística transformandose en el usual estimador no paramétrico, mientras que cuando dicho conocimiento es óptimo $(c \to \infty)$ nuestro estimador se convierte en el mejor que puede ser, en $S_o(t)$, siendo en ambas situaciones el riesgo cero por estar seguros en ambas circunstancias, en una de que nuestro conocimiento a priori era pésimo y en otra que era óptimo.

Comparación de la regla obtenida con la obtenida por Ferguson:

Empecemos escribiendo la regla Bayes obtenida por Ferguson (buscada dentro del conjunto de todas las reglas de decisión) con la obtenida aquí (buscada dentro del conjunto de combinaciones lineales de $S_n(t)$ y $S_o(t)$), ambas en el caso de muestras de tamaño 1.

La de Ferguson es, con nuestra notación

$$E[S(t)/X=x] = \begin{cases} [S_o(t)]^{m(c)} & \text{si} \quad t < x \\ S_o(t)\,[S_o(x)]^{-\frac{1}{1+c}\,m(c)} & \text{si} \quad t \geq x \end{cases}$$

mientras que la nuestra es

$$\hat{S}(t) = \begin{cases} [S_o(t)]^{m(c)} & \text{si} \quad t \leq x \\ S_o(t)\,\dfrac{1 - [S_o(t)]^{m(c)}}{1 - S_o(t)} & \text{si} \quad t > x \end{cases}$$

que como se ve coinciden para $t < x$, y aunque ya vimos al final de la sección III del capítulo 1 que la obtenida por Ferguson representaba la curva de regresión y la nuestra la recta de regresión, con lo que de pérdida de precisión ello conlleva, vamos a apuntar, sin embargo una serie de ventajas de nuestro estimador lineal $\hat{S}(t)$ en relación con el de Ferguson:

(a) Cada regla lineal $\hat{S}(t)$ lleva asociado su correspondiente riesgo Bayes mínimo, mientras que con el tan complicado estimador obtenido por Ferguson, ésto resulta inviable.

(b) En nuestro estimador, parámetros como $c, \Lambda^*(t)$, $m(c)$, etc., tiene una clara interpretación que ya hemos expuesto, mientras que con el estimador $E[S(t)/X=x]$ ninguna interpretación es posible.

(c) En $E[S(t)/X=x]$, el tamaño esperado a posteriori del salto en una observación, depende de donde ocurra la observación, a través de $S_o(x)$. Aquí no ocurre eso, con lo que cabría pensar en algún valor de c como "tamaño muestral a priori".

(d) $\hat{S}(t) \xrightarrow[c \to o]{} S_n(t)$ que es el estimador de máxima verosimilitud, no importando cual fuera el valor escogido para $S_o(t)$, cosa que no ocurre en $E[S(t)/X=x]$.

<u>Tasa de convergencia del riesgo</u>:

Vimos en el teorema 3.2.4. que

$$R_{min}(n) = ([S_o(t)]^{m(c)+1} - [S_o(t)]^2)\ q_n(t)$$

y como $R_{min}(0) = [S_o(t)]^{m(c)+1} - [S_o(t)]^2$, teorema 3.2.1, será

$$R_{min}(n) = q_n(t) \cdot R_{min}(0)$$

y al ser $q_n(t) \leq 1$ como facilmente se puede deducir, convendrá tomar muestra y no considerar el problema sin muestra en orden a disminuir el riesgo en las estimaciones.

Por otro lado, en cuanto a la tasa de convergencia, la cual tiende hacia uno cuando n va a $+\infty$ es, como facilmente puede deducirse,

$$\frac{R_{min}(n)}{R_{min}(n-1)} = \frac{n-1}{n} + \frac{q_n(t)}{n} = 1 - \frac{1}{n}(1 - q_n(t))$$

III. ESTIMACION DE LA FUNCION DE SUPERVIVENCIA CUANDO HAY DATOS CENSURADOS.

Si suponemos el esquema habitual de éste capítulo, de ser $T \geq 0$ una variable aleatoria con función de distribución aleatoria $F(t)$ un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$, la novedad estudiada en ésta sección es la de estimar la función de supervivencia $S(t) = 1 - F(t)$ cuando alguno o algunos de los elementos en la muestra ha sido censurado, es decir ha dejado de existir o de funcionar por alguna causa distinta a la que estamos estudiando. El razonamiento de excluir de la muestra a dichos elementos disminuyendo el tamaño muestral es incorrecto, por poder llegar a perder la objetividad que todo razonamiento científico debe de tener y además por perder información: la de que los individuos censurados tienen un tiempo de fallo mayor que el tiempo de censura. Por supuesto, si consideramos el problema sin muestra, no cabe hablar de datos censurados, ya habiendo sido tratado dicho supuesto en el teorema 3.2.1.

Formalizando las ideas antes expuestas, sea δ el número de individuos censurados en una muestra de tamaño n. Si $S_n(t)$ es como siempre la función de supervivencia muestral, nos proponemos encontrar la regla Bayes para $S(t)$ dentro de la clase de reglas de decisión de la forma

$$a_1 [\bar{S}_n(t)]^{1-\frac{\delta}{n}} + a_2 [\bar{S}_n(t)]^{\frac{\delta}{n}} + b\, S_o(t)$$

en donde a_1, a_2 y b son constantes a determinar, y que claramente es una generalización del estudiado en la sección anterior, ya que cuando $\delta = 0$ se transforma en el allí considerado.

Estableceremos una hipótesis adicional al modelo: supondremos que δ es independiente de T con media en el muestreo conocida $\delta_o = E[\delta]$.

Teorema 3.3.1. Sea $T \geq 0$ una variable aleatoria observable real valorada, la cual supondremos nos indica el tiempo de fallo de un sistema, con función de supervivencia desconocida $S(t)$. Sea P la distribución de probabilidad a priori sobre F, inducida por un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$, y consecuentemente sea $S_o(t) = e^{-\Lambda^*(t)}$ la función de supervivencia a priori. Sea $T_1,\ldots,T_n$ una muestra aleatoria simple de tamaño n de T, en la cual hay δ individuos censurados, con δ independiente de T y con $E[\delta] = \delta_o$ un número conocido.

Entonces, la regla Bayes para $S(t)$, bajo pérdida cuadrática y buscada dicha regla dentro de la clase de reglas de decisión de la forma

$$a_1 [S_n(t)]^{\frac{n-\delta}{n}} + a_2 [S_n(t)]^{\frac{\delta}{n}} + b\, S_o(t)$$

en donde $S_n(t)$ es la función de supervivencia muestral, es

$$\hat{a}_1 [S_n(t)]^{\frac{n-\delta}{n}} + \hat{a}_2 [S_n(t)]^{\frac{\delta}{n}} + \hat{b}\, e^{-\Lambda^*(t)}$$

siendo $(\hat{a}_1, \hat{a}_2, \hat{b})$ el vector dado por la expresión

$$K(\hat{a}_1, \hat{a}_2, \hat{b})' = \vec{L}$$

siendo

$$K = \begin{bmatrix} -1 + \frac{2}{n}\delta_o + \left(2 - \frac{2\delta_o}{n}\right) e^{-\Lambda^*(t)} & e^{-\Lambda^*(t)} & \frac{\delta_o}{n} e^{-\Lambda^*(t)} + \left(1 - \frac{\delta_o}{n}\right) e^{-2\Lambda^*(t)} \\ e^{-\Lambda^*(t)} & 1 - \frac{2}{n}\delta_o + \frac{2}{n}\delta_o e^{-\Lambda^*(t)} & \left(1 - \frac{\delta_o}{n}\right) e^{-\Lambda^*(t)} + \frac{\delta_o}{n} e^{-2\Lambda^*(t)} \\ \frac{\delta_o}{n} e^{-\Lambda^*(t)} + \left(1 - \frac{\delta_o}{n}\right) e^{-2\Lambda^*(t)} & \left(1 - \frac{\delta_o}{n}\right) e^{-\Lambda^*(t)} + \frac{\delta_o}{n} e^{-2\Lambda^*(t)} & e^{-2\Lambda^*(t)} \end{bmatrix}$$

y

$$\vec{L} = \left(\frac{\delta_o}{n} e^{-\Lambda^*(t)} + \left(1 - \frac{\delta_o}{n}\right) e^{-\frac{2c+1}{c+1}\Lambda^*(t)},\ \left(1 - \frac{\delta_o}{n}\right) e^{-\Lambda^*(t)} + \right.$$

$$\left. + \frac{\delta_o}{n} e^{-\frac{2c+1}{c+1}\Lambda^*(t)},\ e^{-2\Lambda^*(t)}\right)$$

Así mismo, el riesgo Bayes asociado a P es

$$R_{min}(n) = e^{-\frac{2c+1}{c+1}\Lambda^*(t)} - \vec{L}'(\hat{a}_1, \hat{a}_2, \hat{b})'$$ "

En efecto:

Como sabemos, es

$$E\left[S(t)\right] = e^{-\Lambda^*(t)} \qquad \text{y}$$

$$E\left[S^2(t)\right] = \left[e^{-\Lambda^*(t)}\right]^{m(c)+1} = e^{-\frac{2c+1}{c+1}\Lambda^*(t)}$$

y además, supondremos como siempre $S_o(t) \neq 0$ y $S_o(t) \neq 1$. El resultado se sigue de la aplicación inmediata del teorema 1.3.12.

IV. ESTIMACION DEL TIEMPO MEDIO DE SUPERVIVENCIA.

Esta sección será muy concreta en resultados y breve en sus razonamientos. Esto es debido a que en la sección IV del capítulo 2 ya se trató también éste tema y no queremos parecer reiterativos ni demasiado recargantes. Quisiera sin embargo justificar la existencia de ésta sección así como de otras que aparecen en éste trabajo dada su clara independencia de la ya mencionada.

Supondremos a lo largo de toda ésta sección las siguientes hipótesis:

(a) $T \geq 0$ es una variable aleatoria con función de distribución desconocida $F(t)$.

(b) $\mathcal{P}$ es la probabilidad a priori definida sobre $\mathcal{F}$, que supondremos inducida por un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$.

(c) Consideraremos siempre la pérdida cuadrática

$$L(g(F),d) = (g(F) - d)^2$$

(d) Calcularemos, mediante dos aproximaciones, reglas Bayes para el tiempo medio de supervivencia, supuesto existente $\mu(F)$, definido en 2.4.1.

En éstas condiciones, la función de supervivencia a priori será

$$S_o(t) = e^{-\Lambda^*(t)} = E[S(t)]$$

$$[S_o(t)]^{m(c)} = e^{-\frac{c}{c+1}\Lambda^*(t)}$$

y la función de distribución a priori

$$F_o(t) = 1 - e^{-\Lambda^*(t)}$$

siendo sus momentos respecto al origen,

$$\mu_o^i = \int_o^\infty t^i \, d\, F_o(t) = -\int_o^\infty t^i \, d\, S_o(t) = \int_F \int_o^\infty t^i \, dF(t) dP(F)$$

que supondremos existentes, por lo que equivalentemente existirán los de $F(t)$.

En particular, el tiempo medio de supervivencia a priori será,

$$\mu_o^1 = \int_o^\infty t \, dF_o(t) = -\int_o^\infty t \, dS_o(t) = \int_F \int_o^\infty t \, dF(t) dP(F).$$

Empecemos por el problema sin muestra.

<u>Teorema 3.4.1</u>. El estimador Bayes para $\mu(F)$ (acción Bayes), respecto a la pérdida cuadrática, en el problema sin muestra, es el tiempo medio de supervivencia a priori μ_o^1, siendo el riesgo Bayes mínimo alcanzado,

$$R_{min}(0) = \int_0^{\infty} S_o(x)\left(\int_0^x [S_o(y)]^{m(c)}\, dy\right)dx +$$

$$+ \int_0^{\infty} [S_o(x)]^{m(c)} \left(\int_x^{\infty} S_o(y)dy\right)dx - (\mu_o^1)^2 =$$

$$= \int_0^{\infty} e^{-\Lambda^*(x)} \left(\int_0^x e^{-\frac{c}{c+1}\Lambda^*(y)}\, dy\right)dx +$$

$$+ \int_0^{\infty} e^{-\frac{c}{c+1}\Lambda^*(x)} \left(\int_x^{\infty} e^{-\Lambda^*(y)}\, dy\right)dx - (\mu_o^1)^2 .$$

Estimación primera:

Si $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} T_i$ es la media muestral, vamos a calcular aquí la regla Bayes para $\mu(F)$ dentro de la clase de reglas de decisión de la forma

$$a\,\bar{x} + b\,\mu_o^1$$

Teorema 3.4.2. En las condiciones expuestas, la regla Bayes para $\mu(F)$ es

$$\hat{\mu}(S) = \frac{n\,K_{11} - n\,(\mu_o^1)^2}{\mu_o^2 + (n-1)K_{11} - n(\mu_o^1)^2}\,\bar{x} + \frac{\mu_o^2 - K_{11}}{\mu_o^2 + (n-1)K_{11} - n(\mu_o^1)^2}\,\mu_o^1$$

y el riesgo Bayes asociado a P,

$$R_{min}(n) = \frac{[K_{11} - (\mu_o^1)^2]\,.\,[\mu_o^2 - K_{11}]}{\mu_o^2 + (n-1)K_{11} - n\,(\mu_o^1)^2}$$

siendo

$$K_{11} = \int_0^{\infty} S_o(x)\left(\int_0^x [S_o(y)]^{m(c)}\, dy\right)dx +$$

$$+ \int_0^\infty [S_o(x)]^{m(c)} \left(\int_x^\infty S_o(y)dy\right)dx =$$

$$= \int_0^\infty e^{-\Lambda^*(x)} \left(\int_0^x e^{-\frac{c}{c+1}\Lambda^*(y)} dy\right)dx +$$

$$+ \int_0^\infty e^{-\frac{c}{c+1}\Lambda^*(x)} \left(\int_x^\infty e^{-\Lambda^*(y)} dy\right)dx$$

Obsérvese que los coeficientes de $\bar{x}$ y μ_o^1 estimados suman uno de acuerdo con el teorema 2.1.1.

Corolario 3.4.1. El estimador $\hat{\mu}(S) = \hat{a}\,\bar{x} + \hat{b}\,\mu_o^1$ converge cuando n tiende hacia ∞, hacia $\mu(S)$ casi seguro. Es decir, el estimador $\hat{\mu}(S)$ es consistente para $\mu(S)$ dentro de la clase de distribuciones con primer momento finito.

Así pues, cuando $n \to \infty$ la única información es la contenida en la muestra ya que el conocimiento a priori no se pondera.

Corolario 3.4.2. El riesgo mínimo asociado a $\hat{\mu}(S)$ converge hacia cero cuando $n \to \infty$.

Como ya indicamos en su momento, la regla Bayes para $\mu(F)$ dentro de la clase de reglas de decisión de la forma

$$a_1 T_1 + \ldots + a_n T_n + b\,\mu_o^1$$

en donde $(T_1,\ldots,T_n)$ es una muestra aleatoria simple, es la misma que la obtenida por el método acabado de exponer, Goldstein (1975).

Estimación segunda:

Ahora supondremos una muestra de tamaño uno y buscaremos la regla Bayes dentro de la clase de reglas de decisión de la forma

$$\lambda_o + \lambda_1 x + \lambda_2 x^2 + \dots + \lambda_n x^n = (1,x,\dots,x^n)\vec{\lambda}_n$$

en donde x es el valor de la variable T en la muestra de tamaño uno.

Teorema 3.4.3. La regla Bayes $\hat{\mu}(S)$ para $\mu(S)$ dentro de la clase de reglas de decisión de la forma

$$(1,x,\dots,x^n)\vec{\lambda}_n = \lambda_o + \lambda_1 x + \dots + \lambda_n x^n$$

es

$$\hat{\mu}_{P_o}(S) = (1, x, x^2,\dots,x^n)\hat{\vec{\lambda}}_n$$

siendo el riesgo Bayes asociado a P

$$R_{min}(n) = K_{11} - (K_{10}, K_{11}, \dots, K_{1n})\hat{\vec{\lambda}}_n$$

en donde

$$\hat{\vec{\lambda}}_n = \begin{pmatrix} \hat{\lambda}_o \\ \hat{\lambda}_1 \\ \vdots \\ \hat{\lambda}_n \end{pmatrix} = \begin{pmatrix} 1 & \mu_o^1 & \dots & \mu_o^n \\ \mu_o^1 & \mu_o^2 & & \mu_o^{n+1} \\ \dots & \dots & \dots & \dots \\ \mu_o^{n+1} & \mu_o^{n+2} & \dots & \mu_o^{2n} \end{pmatrix}^{-1} \cdot \begin{pmatrix} K_{10} \\ K_{11} \\ \vdots \\ K_{1n} \end{pmatrix}$$

supuesto que la función de distribución a priori $F_o(t)$ tiene más de n puntos de incremento y siendo

$$K_{10} = \mu_o^1$$

$$\text{y} \quad K_{ij} = K_{ji} = i.j \int_0^\infty x^{i-1}\, S_o(x)\left(\int_0^x y^{j-1}\, [S_o(y)]^{m(c)}\, dy\right)dx +$$

$$+ \; i.j \int_0^\infty x^{i-1}\, [S_o(x)]^{m(c)} \,.\, \left(\int_x^\infty y^{j-1}\, S_o(y)dy\right)dx =$$

$$= i.j \int_0^\infty x^{i-1}\, e^{-\Lambda^*(t)} \left(\int_0^x y^{j-1}\, e^{-\frac{c}{c+1}\Lambda^*(y)}\, dy\right)dx +$$

$$+ \; i.j \int_0^\infty x^{i-1}\, e^{-\frac{c}{c+1}\Lambda^*(x)} \left(\int_x^\infty y^{j-1}\, e^{-\Lambda^*(y)}\, dy\right)dx$$

si $i \geq 1$, $j \geq 1$.

Corolario 3.4.3. La relación entre $R_{min}(n-1)$ y $R_{min}(n)$ es

$$R_{min}(n) - R_{min}(n-1) = (\hat{\lambda}_n^*)^2 \; \frac{|D_n|}{|D_{n-1}|}$$

en donde $\hat{\lambda}_n^*$ es el coeficiente estimado de x^n, dentro de la clase de reglas de decisión que son polinomios de grado n.

V. ESTIMACION DEL MOMENTO DE ORDEN i.

La hipótesis aquí serán las mismas que en la sección anterior:

(a) $T \geq 0$ es una variable aleatoria con función de distribución desconocida $F(t)$.

(b) La probabilidad a priori P, es la inducida por un proceso homogéneo simple de parámetros $(\Lambda^*(t), c)$.

(c) La función de pérdida es cuadrática

$$L(g(F),\, d) = (g(F) - d)^2$$

(d) Pretendemos calcular reglas Bayes por dos métodos diferentes " para el momento de orden i

$$\mu_i(F) = \int_0^\infty t^i \, dF(t) = -\int_0^\infty t^i \, dS(t) = i \int_0^\infty t^{i-1} S(t)dt = \mu_i(S)$$

Teorema 3.5.1. El estimador Bayes para $\mu_i(F)$ (acción Bayes) en el problema sin muestra es el momento de orden i respecto al origen de $F_o(t)$, μ_o^i, siendo el riesgo Bayes mínimo asociado

$$R_{min}(0) = K_{ii} - (\mu_o^i)^2$$

en donde K_{ii} es el definido en la sección anterior.

Estimación primera:

Si ahora buscamos la regla Bayes para $\mu_i(F)$ dentro de la clase de reglas de decisión de la forma

$$b_o \, \mu_o^i + b_1 \, \overline{\mu}_n^1 + b_2 \, \overline{\mu}_n^2 + \ldots + b_i \, \overline{\mu}_n^i$$

en donde

$$\overline{\mu}_n^j = \frac{1}{n} \sum_{h=1}^{n} (x_h)^j, \qquad j=1,2,\ldots,i$$

es el momento muestral de orden j y si como antes

$$K_{ji} = K_{ij} = i.j \int_0^\infty x^{i-1} \, e^{-\Lambda^*(x)} \left(\int_0^x y^{j-1} \, e^{-\frac{c}{c+1}\Lambda^*(t)} dy\right)dx +$$

$$+ \, i.j \int_0^\infty x^{i-1} \, e^{-\frac{c}{c+1}\Lambda^*(x)} \left(\int_x^\infty y^{j-1} \, e^{-\Lambda^*(x)} dx\right)dy \qquad j=1,\ldots,i$$

tenemos el siguiente resultado.

Teorema 3.5.2. La regla Bayes $\hat{\mu}_i(F)$ para $\mu_i(F)$ dentro de la clase de estimadores de la forma

$$(\mu_o^i, \bar{\mu}_n^1, \bar{\mu}_n^2, \dots, \mu_n^i)\vec{b}_i = b_o\, \mu_o^i + b_1\, \bar{\mu}_n^1 + b_2\, \bar{\mu}_n^2 + \dots + b_i\, \bar{\mu}_n^i$$

es

$$\hat{\mu}_i(F) = (\mu_o^i, \bar{\mu}_n^1, \bar{\mu}_n^2, \dots, \bar{\mu}_n^i)\hat{\vec{b}}_i$$

siendo el riesgo Bayes asociado a P

$$R_{min}(n) = K_{ii} - ((\mu_o^i)^2, K_{i1}, K_{i2}, \dots, K_{ii})\, \hat{\vec{b}}_i$$

en donde

$$\hat{\vec{b}}_i = \begin{bmatrix} \hat{b}_o \\ \hat{b}_1 \\ \vdots \\ \hat{b}_n \end{bmatrix} = \begin{bmatrix} (\mu_o^i)^2 & \mu_o^i \mu_o^1 & \dots & (\mu_o^i)^2 \\ \mu_o^i \mu_o^1 & \frac{1}{n}\mu_o^2 + \frac{n-1}{n} K_{11} & \dots & \frac{1}{n}\mu_o^{i+1} + \frac{n-1}{n} K_{1i} \\ \dots & \dots & \dots & \dots \\ (\mu_o^i)^2 & \frac{1}{n}\mu_o^{i+1} + \frac{n-1}{n} K_{i1} & \dots & \frac{1}{n}\mu_o^{2i} + \frac{n-1}{n} K_{ii} \end{bmatrix}^{-1} \begin{bmatrix} (\mu_o^i)^2 \\ K_{i1} \\ \vdots \\ K_{ii} \end{bmatrix}$$

<u>Corolario 3.5.1</u>. Cuando $n \to \infty$ los coeficientes $\hat{b}_o, \hat{b}_1, \dots, \hat{b}_{i-1}$ convergen hacia o y $\hat{b}_i$ hacia 1, es decir el vector $\hat{\vec{b}}_i$ converge hacia el vector $(0,0,\dots,0,1)$. Así pues, cuando $n \to \infty$, la única información es la contenida en la muestra, en concreto en el momento muestral de orden i.

Además $R_{min}(n) \longrightarrow 0$, cuando $n \to \infty$.

Estimación segunda:

Los vamos a restringir ahora en la búsqueda de la regla Bayes para $\mu_i(F)$ a la clase de reglas de decisión de la forma

$$\beta_o + \beta_1 x + \ldots + \beta_n x^n = (1, x, \ldots, x^n)\vec{\beta}_n$$

en donde x es el valor de la variable aleatoria T en la muestra de tamaño 1.

Supondremos que $F_o(t)$ tiene más de n puntos de incremento para que el determinante de la matriz a invertir más abajo sea distinto de cero.

Teorema 3.5.3. La regla Bayes para $\mu_i(S)$ dentro de la clase de reglas de decisión de la forma

$$(1, x, x^2, \ldots, x^n)\vec{\beta}_n = \beta_o + \beta_1 x + \ldots + \beta_n x^n$$

es

$$(1, x, x^2, \ldots, x^n)\hat{\vec{\beta}}_n = \hat{\beta}_o + \hat{\beta}_1 x + \ldots + \hat{\beta}_n x^n$$

siendo el riesgo Bayes asociado a P

$$R_{min}(n) = K_{ii} - (K_{io}, K_{i1}, \ldots, K_{in})\hat{\vec{\beta}}_n$$

en donde

$$\hat{\vec{\beta}}_n = \begin{pmatrix} \hat{\beta}_o \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_n \end{pmatrix} = \begin{pmatrix} 1 & \mu_o^1 & \cdots & \mu_o^n \\ \mu_o^1 & \mu_o^2 & \cdots & \mu_o^{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_o^{n+1} & \mu_o^{n+2} & \cdots & \mu_o^{2n} \end{pmatrix}^{-1} \begin{pmatrix} K_{io} \\ K_{i1} \\ \vdots \\ K_{in} \end{pmatrix}$$

Corolario 3.5.2. La relación entre $R_{min}(n-1)$ y $R_{min}(n)$ es

$$R_{min}(n) - R_{min}(n-1) = (\hat{\beta}_n^*)^2 \frac{|D_n|}{|D_{n-1}|}$$

en donde $\hat{\beta}_n^*$ es el coeficiente estimado para x^n, dentro de la clase de reglas de decisión que son polinomios de grado n.

VI. ESTIMACION DE LA FUNCION DE DISTRIBUCION

Supondremos la siguiente situación:

(a) $T \geq 0$ es una variable aleatoria con función de distribución desconocida $F(t)$.

(b) La probabilidad a priori P es la inducida por un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$.

(c) La función de pérdida es cuadrática

$$L(g(F), d) = (g(F) - d)^2$$

(d) Pretendemos calcular la regla Bayes para la función de distribución en un punto fijo $t \geq 0$. Es decir, $g(F) = F(t)$.

Ya que hemos determinado la regla Bayes para $S(t)$ en la sección II y $F(t) = 1 - S(t)$, podemos hacer uso del teorema 2.1.2 ya que la función

$$f(S(t)) = 1 - S(t)$$

es una función lineal, con lo que un estimador Bayes para $F(t)$ será

$$\hat{F}(t) = f(\hat{S}(t)) = 1 - \hat{S}(t)$$

es decir,

$$\hat{F}(t) = 1 - [1-q_n(t)]\,(1-F_n(t)) - q_n(t)(1-F_o(t)) =$$

$$= [1-q_n(t)]\,F_n(t) + q_n(t)\,F_o(t)$$

siendo el riesgo Bayes de $\mathcal{P}$ el mismo que en la estimación de $S(t)$.

Podemos pues enunciar dos teoremas, el primero para el problema sin muestra y el segundo cuando hay muestra, siendo en éste último caso la regla Bayes la antes determinada, y buscada dentro del conjunto de reglas de decisión de la forma

$$f(a\,S_n(t) + b\,S_o(t)) = 1 - a\,S_n(t) - b\,S_o(t) =$$

$$= a\,F_n(t) + b\,F_o(t)$$

<u>Teorema 3.6.1</u>. En las condiciones antes expuestas, el estimador Bayes $\hat{F}(t)$, para $F(t)$ en el problema sin muestra es la función de distribución a priori $F_o(t) = 1 - e^{-\Lambda^*(t)}$, siendo el riesgo mínimo asociado

$$R_{min}(0) = [S_o(t)]^{m(c)+1} - [S_o(t)]^2 = e^{-\frac{2c+1}{c+1}\Lambda^*(t)} - e^{-2\Lambda^*(t)}$$

El resultado se sigue de ser $E[F(t)] = F_o(t)$ el que hace mínimo el riesgo Bayes

$$\int_F (F(t) - d)^2\, d\mathcal{P}(F)$$

y de ser $R_{min}(0) = V(F(t)) = V(S(t))$.

<u>Teorema 3.6.2</u>. En el problema con muestras, si n es el tamaño muestral en la última etapa, el estimador Bayes $\hat{F}(t)$ para $F(t)$, en la situación aquí considerada es

$$\hat{F}(t) = \left[1 - q_n(t)\right] F_n(t) + q_n(t) F_o(t)$$

siendo el riesgo Bayes de $\mathcal{P}$

$$R_{min}(n) = \frac{-\left[S_o(t)\right]^2 + \left[S_o(t)\right]^{m(c)+2} + \left[S_o(t)\right]^{m(c)+1} - \left[S_o(t)\right]^{2m(c)+1}}{1 + (n-1)\left[S_o(t)\right]^{m(c)} - n\left[S_o(t)\right]}$$

en donde

$$S_o(t) = e^{-\Lambda^*(t)}$$

$$q_n(t) = \frac{1 - \left[S_o(t)\right]^{m(c)}}{1+(n-1)\left[S_o(t)\right]^{m(c)} - n\left[S_o(t)\right]}$$

y $$m(c) = \frac{c}{c+1} .$$

<u>Comportamiento asintótico muestral</u>:

<u>Teorema 3.6.3</u>. En el esquema aquí considerado de ser $\mathcal{P}$ la inducida por un proceso homogéneo simple de parámetros $(\Lambda^*(t),c)$, el estimador

$$\hat{F}(t) = \left[1 - q_n(t)\right] F_n(t) + q_n(t) F_o(t)$$

verifica:

(a) $\hat{F}(t)$ converge casi seguro hacia $F(t)$ cuando $n \to \infty$ para cada $t \geq 0$ fijo.

(b) $\hat{F}(t)$ converge en probabilidad y en ley hacia $F(t)$ y la función característica asociada a $\hat{F}(t)$ hacia la asociada a $F(t)$, cuando $n \to \infty$, para cada $t \geq 0$.

(c) $\hat{F}(t)$ es un estimador consistente para $F(t)$.

Los resultados se siguen de la misma manera que en la sección VI del capítulo anterior.

Comportamiento de $\hat{F}(t)$ en función de c:

A partir del teorema 3.2.4 obtenemos el siguiente resultado

Teorema 3.6.4. El estimador Bayes $\hat{F}(t)$ y el riesgo mínimo asociado determinados en el teorema 3.6.2 verifican que

(a) $\lim_{c \to o} \hat{F}(t) = F_n(t)$ y $\lim_{c \to o} R_{min}(n) = 0$

(b) $\lim_{c \to \infty} \hat{F}(t) = 1-e^{-\Lambda^*(t)} = F_o(t)$ y $\lim_{c \to \infty} R_{min}(n) = 0$

Tasa de convergencia del riesgo:

A partir del teorema 3.2.4. obtenemos que la tasa de convergencia del riesgo, la cual tiende hacia uno al tender n hacia $+\infty$ es

$$\frac{R_{min}(n)}{R_{min}(n-1)} = \frac{n-1}{n} + \frac{q_n(t)}{n}$$

VII. PROBLEMA DE LAS DOS MUESTRAS

Si suponemos ahora que $F(x)$ es una función de distribución aleatoria, un proceso homogéneo simple de parámetros $(\Lambda^*(x),c)$ y $G(y)$ otra distribución aleatoria, ésta un proceso homogéneo simple de parámetros $(\Lambda'(y),c)$ y si P y P' son respectivamente las probabilidades inducidas por los procesos, las cuales supondremos independientes, al tratar de estimar

$$\Delta = \int_0^\infty F(y)\, dG(y)$$

bajo pérdida cuadrática obtenemos el siguiente resultado.

Teorema 3.7.1. Si $X_1,\ldots,X_m$ es una muestra extraida mediante F e $Y_1,\ldots,Y_n$ una muestra extraida mediante G y si

$$F_o(t) = 1 - e^{-\Lambda^*(t)}$$

y

$$G_o(t) = 1 - e^{-\Lambda'(t)}$$

son las respectivas funciones de distribución a priori, la regla Bayes para Δ es

$$\hat{\Delta}_{m,n} = \int \hat{F}(t)\, d\hat{G}(t) = q_m(t)\, q_n'(t)\Delta_o + q_m(t)\,(1-q_n'(t)).$$

$$.\ \frac{1}{n}\sum_{j=1}^{n} F_o(Y_j) + (1-q_m(t))\, q_n'(t)\,\frac{1}{m}\sum_{i=1}^{m}\left[1 - G_o(X_i^-)\right] +$$

$$+ (1 - q_m(t))(1 - q_n'(t))\,\frac{U}{n.m}$$

en donde

$$\Delta_o = \int_0^\infty F_o(y)dG_o(y)$$

$$q_m(t) = \frac{1 - [S_o(t)]^{\frac{c}{c+1}}}{1+(m-1)\,[S_o(t)]^{\frac{c}{c+1}} - m\,S_o(t)}; \qquad S_o(t) = 1 - F_o(t)$$

$$q_n'(t) = \frac{1 - [S_o'(t)]^{\frac{c'}{c'+1}}}{1+(n-1)\,[S_o'(t)]^{\frac{c'}{c'+1}} - n\,S_o'(t)}; \qquad S_o'(t) = 1 - G_o(t)$$

U el número de pares (X_i,Y_j) para los cuales $X_i \leq X_j$, y buscada dicha regla dentro del conjunto de reglas de decisión de la forma

$$a_1\,\Delta_o + a_2\,\frac{1}{n}\sum_{j=1}^{n} F_o(Y_j) + a_3\,\frac{1}{m}\sum_{i=1}^{m}\left[1 - G_o(X_i^-)\right] + a_4\,\frac{U}{n.m}$$

"

El resultado se sigue facilmente del teorema 2.1.3 por ser la función

$$f(x,y) = \int_0^\infty (1-x)d(1-y)$$

bilineal.

<u>Comportamiento asintótico muestral</u>:

Dado que

$$\lim_{m\to\infty} q_m(t) = 0 \qquad y \qquad \lim_{n\to\infty} q_n'(t) = 0$$

al tender n y m a $+\infty$, $\hat{\Delta}_{m,n} \longrightarrow \frac{U}{n.m}$, el estimador de Mann-Whitney (1947).

<u>Comportamiento asintótico en función de c y c'</u>:

Dado que

$$\lim_{c\to o} q_m(t) = 0 \qquad y \qquad \lim_{c'\to o} q_n'(t) = 0$$

se sigue que

$$\hat{\Delta}_{m,n} \xrightarrow[\substack{c\to o\\ c'\to o}]{} \frac{U}{n.m}$$

el usual estimador no paramétrico.

CAPITULO 4

ESTIMACION CON DISTRIBUCION A PRIORI UN PROCESO GAMMA EXTENDIDO

I. INTRODUCCION

En éste capítulo vamos de nuevo a hacer estimaciones de una función de F, la función de distribución desconocida de una variable aleatoria $T \geq 0$, la cual nos indica el tiempo de fallo de un sistema. Sin embargo, ahora la probabilidad a priori P definida sobre F va a ser la inducida por un proceso gamma extendido, proposición 1.8.2. Obsérvese que a diferencia de los procesos gamma exponenciales y de los homogéneos simples, los gamma extendidos no son neutrales a la derecha, aunque como luego veremos, a la hora de las estimaciones alcanzaremos las mismas cotas que con aquellos, siendo de nuevo la distribución a posteriori poco manejable.

Supondremos durante todo el tiempo que la función de pérdida es cuadrática

$$L(F,d) = [g(F) - d]^2$$

Como sabemos, para inducir la probabilidad a priori P debemos de determinar el proceso gamma extendido, o lo que es lo mismo, aparte del proceso con incrementos independientes $Z(t)$, determinar las funciones $\alpha(t)$ y $\beta(t)$. Una posible via consiste en definir dos funciones media y varianza no decrecientes $\mu(t)$ y $\sigma^2(t)$, y considerar a $\mu(t)$ como la mejor "pista" o conocimiento a priori acerca de la tasa de azar aleatoria $r(t)$ y a $\sigma^2(t)$ como a una medida de la incertidumbre o variación de la tasa de azar aleatoria en el punto t. Supondremos que tanto $\mu(t)$, $\sigma^2(t)$ como $\alpha(t)$ son todas diferenciables, por lo que por la proposición 1.8.3, será

$$\mu(t) = \int_{[0,t)} \beta(s)\ \alpha'(s)ds$$

y

$$\sigma^2(t) = \int_{[0,t)} \beta^2(s)\ \alpha'(s)ds$$

de donde

$$\beta(t) = \frac{\frac{d\sigma^2(t)}{dt}}{\frac{d\ \mu(t)}{dt}}$$

$$\frac{d\ \alpha(t)}{dt} = \frac{\left[\frac{d\ \mu(t)}{dt}\right]^2}{\frac{d\ \sigma^2(t)}{dt}}$$

y tendremos determinada consecuentemente la distribución a priori P. Así pues, con los procesos gamma extendidos operaremos definiendo las funciones $\mu(t)$ y $\sigma^2(t)$, que juegan el mismo papel que jugaban $\Lambda^*(t)$ y c en los gamma exponenciales y en los homogéneos simples, ya que $\sigma^2(t)$ varía entre 0 y $+\infty$, siendo cuando $\sigma^2(t) \to 0$, $r(t)$ degenerada en $\mu(t)$ y nuestra estimación a priori $\mu(t)$ sobre $r(t)$ será óptima, mientras que cuando $\sigma^2(t) \to \infty$, $V(r(t)) \to \infty$ y $\mu(t)$ será una pésima estimación a priori sobre $r(t)$.

Hay que hacer notar también que los trascendentes teoremas 2.1.1, 2.1.2 y 2.1.3 siguen siendo válidos aquí.

Así pues, las hipótesis que supondremos durante todo éste capítulo y que no repetiremos en cada sección para no ser reiterativos son:

(a) P, la probabilidad a priori sobre $\mathcal{F}$, es la inducida por un proceso gamma extendido de parámetros $(\alpha(.),\ \beta(.))$.

(b) La función de pérdida es cuadrática,

$$L(F,d) = (g(F) - d)^2$$

(c) Trataremos de estimar funciones $g(F)$, en donde g es una función definida sobre $\mathcal{F}$, medible.

(d) Nos restringiremos en la estimación de $g(F)$ a la clase de reglas de decisión que son combinaciones lineales de ciertas funciones de la muestra que especificaremos en cada caso.

II. <u>ESTIMACION DE LA FUNCION DE SUPERVIVENCIA CUANDO NO HAY DATOS CENSURADOS</u>.

Supondremos aquí que $g(F) = S(t) = 1 - F(t)$ y que la clase de reglas de decisión en la cual buscamos la regla Bayes es la expresada en el teorema 1.3.9.

Dado que

$$E[S(t)] = \exp\left[-\int_{[0,t)} \log\,(1+\beta(s)(t-s))d\alpha(s)\right] = S_0(t)$$

por la proposición 1.8.4, y que

$$E[S^2(t)] = \exp\left[-\int_{[0,t)} \log\,(1+2\beta(s)(t-s))d\alpha(s)\right]$$

por el teorema 1.8.1, podemos enunciar dos teoremas, uno para el problema sin muestra y otro cuando las haya.

<u>Teorema 4.2.1</u>. El estimador Bayes para $S(t)$ en el problema sin muestra es la función de supervivencia a priori $S_0(t)$, siendo el riesgo mínimo asociado,

$$R_{min}(0) = \exp\left[-\int_{[0,t)} \log(1+2\beta(s)(t-s))\,d\alpha(s)\right] -$$

$$- \exp\left[-2\int_{[0,t)} \log\,(1+\beta(s)(t-s))\,d\alpha(s)\right]$$

La demostración se sigue de alcanzarse el valor mínimo del riesgo para $S_o(t) = E[S(t)]$, y de ser el riesgo mínimo asociado $R_{min}(0) = V(S(t))$.

En cuanto al problema con muestras, como aplicación inmediata del teorema 1.3.9, obtenemos el siguiente resultado.

Teorema 4.2.2. La regla Bayes $\hat{S}(t)$ para $S(t)$ es

$$\hat{S}(t) = [1-W_n(t)]\; S_n(t) + W_n(t)\; \exp\left[-\int_{[0,t)} \log\,(1+\beta(s)(t-s))\,d\alpha(s)\right]$$

siendo el riesgo Bayes de P,

$$R_{min}(n) = R_{min}(0)\;.\;W_n(t)$$

en donde

$$W_n(t) = \frac{1 - \exp\left[-\int_{[0,t)} \log\left(\frac{1+2\beta(s)(t-s)}{1+\beta(s)(t-s)}\right) d\alpha(s)\right]}{1+(n-1)\exp\left[-\int_{[0,t)} \log\left(\frac{1+2\beta(s)(t-s)}{1+\beta(s)(t-s)}\right) d\alpha(s)\right] - n\cdot\exp\left[-\int_{[0,t)} \log\left(1+\beta(s)(t-s)\right) d\alpha(s)\right]}$$

y $S_n(t)$ es la función de supervivencia muestral.

Comportamiento asintótico muestral:

A partir del teorema 1.3.10 y de los corolarios 1.3.1 y 1.3.2 obtenemos el siguiente resultado.

<u>Teorema 4.2.3</u>. El estimador $\hat{S}(t)$ verifica que

(a) $\hat{S}(t)$ converge casi seguro hacia $S(t)$ cuando $n \to \infty$, para cada $t \geq 0$.

(b) El riesgo Bayes mínimo alcanzado $R_{min}(n)$, converge hacia cero, cuando $n \to \infty$.

(c) $\hat{S}(t)$ converge en probabilidad hacia $S(t)$.

(d) $\hat{S}(t)$ converge en ley hacia $S(t)$.

(e) $\hat{S}(t)$ tiene una función característica que converge hacia la función característica de $S(t)$.

(f) $\hat{S}(t)$ es un estimador consistente para $S(t)$.

Antes de finalizar conviene decir algo acerca de la estimación a priori de $S(t)$ y de su grado de confianza por parte del decisor. Ya apuntábamos la interpretación Bayesiana de los parámetros $\mu(t)$ y $\sigma^2(t)$ como estimación a priori y grado de confianza en dicha estimación al hablar de la tasa de azar; sin embargo, al estimar la función de supervivencia, es claro que $S_o(t)$ funciona como estimación a priori acerca de $S(t)$, aunque para ver claramente el grado de confianza en ella mediante un parámetro, deberemos definir antes $\sigma^2(t)$, ya que en el caso de los procesos gamma exponenciales o de los homogéneos simples el parámetro que lo media era c, una constante y no una función que depende de t como $\sigma^2(t)$.

Tasa de convergencia del riesgo:

Como $W_n(t) \leq 1$, $R_{min}(n) \leq R_{min}(0)$, lo cual nos indica la necesidad de tomar muestras, siendo éstas lo más grande posible, ya que en ese caso, como el teorema 4.2.3 nos dice, $R_{min}(n)$ tenderá hacia cero.

Por otro lado, en cuanto a la tasa de convergencia del riesgo, a partir del teorema 4.2.2 deducimos que ésta será:

$$\frac{R_{min}(n)}{R_{min}(n-1)} = \frac{W_n(t)}{W_{n-1}(t)} = 1 - \frac{1}{n}(1 - W_n(t))$$

la cual tiende a 1 al tender n hacia infinito, por ser

$$\lim_{n \to \infty} W_n(t) = 0$$

III. ESTIMACION DE LA FUNCION DE SUPERVIVENCIA CUANDO HAY DATOS CENSURADOS.

De nuevo $g(F) = S(t) = 1 - F(t)$, aunque ahora admitiremos la posibilidad de un número δ de datos censurados, suponiendo que δ es independiente de T, la variable aleatoria tiempo de fallo, y siendo $\delta_o = E[\delta]$ un número conocido. En éstas circunstancias, podemos aplicar el teorema 1.3.12, buscando nuestro estimador Bayes dentro de la clase de reglas de decisión de la forma

$$a_1[S_n(t)]^{\frac{n-\delta}{n}} + a_2[S_n(t)]^{\frac{\delta}{n}} + a_3 E\, S(t)$$

y obtener el siguiente resultado.

Teorema 4.3.1. La regla Bayes para $S(t)$ en las condiciones mencionadas es

$$\hat{a}_1 \left[S_n(t)\right]^{\frac{n-\delta}{n}} + \hat{a}_2 \left[S_n(t)\right]^{\frac{\delta}{n}} + \hat{a}_3 S_o(t)$$

siendo $(\hat{a}_1, \hat{a}_2, \hat{a}_3)$ el vector dado por

$$K(\hat{a}_1, \hat{a}_2, \hat{a}_3)' = \vec{L}$$

en donde K es la matriz

$$K = \begin{bmatrix} -1+\frac{2}{n}\delta_o+\left(2-\frac{2\delta_o}{n}\right)S_o(t) & S_o(t) & \frac{\delta_o}{n}S_o(t)+\left(1-\frac{\delta_o}{n}\right)\left[S_o(t)\right]^2 \\ S_o(t) & 1-\frac{2}{n}\delta_o+\frac{2}{n}\delta_o S_o(t) & \left(1-\frac{\delta_o}{n}\right)S_o(t)+\frac{\delta_o}{n}\left[S_o(t)\right]^2 \\ \frac{\delta_o}{n}S_o(t)+\left(1-\frac{\delta_o}{n}\right)\left[S_o(t)\right]^2 & \left(1-\frac{\delta_o}{n}\right)S_o(t)+\frac{\delta_o}{n}\left[S_o(t)\right]^2 & \left[S_o(t)\right]^2 \end{bmatrix}$$

y

$$\vec{L} = \begin{bmatrix} \frac{\delta_o}{n} S_o(t) + \left(1-\frac{\delta_o}{n}\right) \exp\left[-\int_{[0,t)} \log\left[1+2\beta(s)(t-s)\right] d\alpha(s)\right] \\ \left(1-\frac{\delta_o}{n}\right) S_o(t) + \frac{\delta_o}{n} \exp\left[-\int_{[0,t)} \log\left[1+2\beta(s)(t-s)\right] d\alpha(s)\right] \\ \left[S_o(t)\right]^2 \end{bmatrix}$$

siendo $S_o(t) = \exp\left[-\int_{[0,t)} \log\left(1+\beta(s)(t-s)\right) d\alpha(s)\right]$.

Así mismo, el riesgo Bayes mínimo asociado a P, es

$$R_{min}(n) = \exp\left[-\int_{[0,t)} \log\,[1+2\beta(s)(t-s)]\,d\alpha(s)\right] - \vec{L}'(\hat{a}_1,\hat{a}_2,\hat{a}_3)$$

IV. ESTIMACION DEL TIEMPO MEDIO DE SUPERVIVENCIA.

Como siempre, el tiempo medio de supervivencia se define por

$$\mu(F) = \int_0^\infty t\,dF(t) = -\int_0^\infty t\,dS(t) = \int_0^\infty S(t)dt = \mu(S)$$

y de lo que trataremos ahora será de encontrar estimadores para él. De nuevo, si $F_o(t)$ es la función de distribución a priori, el momento de orden i respecto al origen será:

$$\mu_o^i = \int_0^\infty t^i\,dF_o(t) = -\int_0^\infty t^i\,dS_o(t) = \int_F\int_0^\infty t^i\,dF(t)dP(F)$$

los cuales supondremos existen, con lo que equivalentemente existirán los de $F(t)$.

En particular, para $i=1$, tenemos lo que llamaremos tiempo medio de supervivencia a priori

$$\mu_o^1 = \int_0^\infty t\,dF_o(t) = -\int_0^\infty t\,dS_o(t) = \int_F\int_0^\infty t\,dF(t)dP(F)$$

Comencemos nuestras estimaciones con el problema sin muestras.

Teorema 4.4.1. El estimador Bayes para $\mu(F)$ en el problema sin muestra es el tiempo medio de supervivencia a priori μ_o^1, siendo el riesgo mínimo asociado,

$$R_{min}(0) = K_{11} - (\mu_o^1)^2$$

en donde K_{11} es especificado más abajo.

El resultado se sigue de ser $E[\mu(F)] = \int_F \int_0^\infty t dF(t) dP(F) = \mu_o^1$ el valor que hace mínimo el riesgo

$$\int_F (\mu(F) - d)^2 \, dP(F)$$

En cuanto al riesgo Bayes de P, o riesgo mínimo alcanzado, este será

$$R_{min}(0) = V(\mu(F)) = E[\mu^2(F)] - (E[\mu(F)])^2 = E[\mu^2(F)] - (\mu_o^1)^2$$

y por el teorema 1.8.7 concluimos que

$$E[\mu^2(F)] = \int_F (\int_0^\infty t \, dF(t))(\int_0^\infty u \, dF(u)) dP(F) =$$

$$= \int_0^\infty (\int_0^x \exp\Big[-\int_{[0,y)} \log\,[1+\beta(s)(y+x-2s)]\,d\alpha(s) -$$

$$- \int_{[y,x)} \log\,[1+\beta(s)(x-s)]\,d\alpha(s)\Big] dy) dx +$$

$$+ \int_0^\infty (\int_x^\infty \exp\Big[-\int_{[0,x)} \log[1+\beta(s)(x+y-2s)]\,d\alpha(s) -$$

$$- \int_{[x,y)} \log\,[1+\beta(s)(y-s)]\,d\alpha(s)\Big] dy) dx$$

que de ahora en adelante llamaremos K_{11}, y generalizando dicha notación, en lo que resta llamaremos K_{ij} $i\geq 1$, $j\geq 1$ a

$$K_{ij} = i.j \int_0^\infty x^{i-1} (\int_0^x \exp\Big[-\int_{[0,y)} \log\,[1+\beta(s)(y+x-2s)]\,d\alpha(s) -$$

$$- \int_{[y,x)} \log\,[1+\beta(s)(x-s)]\,d\alpha(s)\Big] y^{j-1}\, dy)\,.\, dx +$$

$$+ i.j \int_0^\infty x^{i-1} (\int_x^\infty \exp\Big[-\int_{[0,x)} \log\,[1+\beta(s)(x+y-2s)]\,d\alpha(s) -$$

$$- \int_{[x,y)} \log \left[1+\beta(s)(y-s)\right] d\alpha(s) \Big] y^{j-1}\, dy)dx$$

Vamos ahora con el problema con muestras.

Estimación primera:

Si $T_1, \ldots, T_n$ es una muestra de tamaño n y $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} T_i$ es la media muestral, buscaremos la regla Bayes para $\mu(S)$ dentro de la clase de reglas de decisión de la forma

$$a\,\bar{x} + b\,\mu_o^1$$

Teorema 4.4.2. La regla Bayes para $\mu(F)$ en las condiciones hasta aquí expresadas es

$$\hat{\mu}(S) = \frac{n\,K_{11} - n(\mu_o^1)^2}{\mu_o^2 + (n-1)K_{11} - n(\mu_o^1)^2}\,\bar{x} + \frac{\mu_o^2 - K_{11}}{\mu_o^2 + (n-1)K_{11} - n(\mu_o^1)^2}\,\mu_o^1$$

y el riesgo Bayes asociado

$$R_{min}(n) = \frac{\left[K_{11} - (\mu_o^1)^2\right] \cdot \left[\mu_o^2 - K_{11}\right]}{\mu_o^2 + (n-1)K_{11} - n\,(\mu_o^1)^2}$$

El resultado se sigue de forma totalmente análoga a la expuesta en el capítulo de las estimaciones con procesos gamma exponenciales, haciendo mínimo el riesgo Bayes

$$\int_F \int_{[0,\infty)^n} (\mu(S) - a\bar{x} - b\mu_o^1)^2\, dF(\bar{x})\,dP(F)$$

en donde por $dF(\bar{x})$ queremos representar $dF(\bar{x}) = dF(x_1) \ldots dF(x_n)$ y teniendo en cuenta que

$$\iint \mu(S)\ dF(\bar{x})\ dP(F) = \mu_o^1$$

$$\iint \bar{x}\ dF(\bar{x})\ dP(F) = \mu_o^1$$

$$\iint \bar{x}\ \mu(S)\ dF(\bar{x})\ dP(F) = K_{11}$$

$$\iint (\bar{x})^2\ dF(\bar{x})\ dP(F) = \frac{1}{n}\ \mu_o^1 + \frac{n-1}{n}\ K_{11}$$

a partir del teorema 1.8.7.

Por último obsérvese que si expresamos el estimador obtenido por

$$\hat{\mu}(S) = \hat{a}\ \bar{x} + \hat{b}\ \mu_o^1$$

es $\hat{a} + \hat{b} = 1$, como tenía que salir. Además, dado que

$$\lim_{n \to \infty} \hat{b} = 0$$

tenemos el siguiente corolario.

Corolario 4.4.1. El estimador $\hat{\mu}(S) = \hat{a}\ \bar{x} + \hat{b}\ \mu_o^1$ converge casi seguro hacia $\mu(S)$ cuando $n \to \infty$. Es decir, el estimador $\hat{\mu}(S)$ es consistente para $\mu(S)$ dentro de la clase de distribuciones con primer momento finito.

Así pues, al ir n hacia $+\infty$ el conocimiento a priori no se pondera y la única información a priori es la contenida en la muestra.

En cuanto al riesgo, tenemos el siguiente resultado.

Corolario 4.4.2. El riesgo mínimo asociado a $\hat{\mu}(S)$ converge hacia cero cuando n va a $+\infty$.

Estimación segunda:

Consideraremos ahora muestras de tamaño uno, siendo x el valor de esa muestra de tamaño uno de la variable aleatoria T, y buscaremos dicha regla Bayes dentro de la clase de reglas de decisión que son los polinomios de la forma

$$\lambda_o + \lambda_1 x + \lambda_2 x^2 + \ldots + \lambda_n x^n = (1, x, \ldots, x^n)\vec{\lambda}_n$$

Teorema 4.4.3. La regla Bayes $\hat{\mu}_{P_o}(S)$ para $\mu(S)$ dentro de la clase de reglas de decisión de la forma

$$(1, x, \ldots, x^n)\vec{\lambda}_n$$

es

$$\hat{\mu}_{P_o}(S) = \hat{\lambda}_o + \hat{\lambda}_1 x + \ldots + \hat{\lambda}_n x^n = (1, x, \ldots, x^n)\hat{\vec{\lambda}}_n$$

siendo el riesgo Bayes asociado a P,

$$R_{min}(n) = K_{11} - (K_{10}, K_{11}, \ldots, K_{1n})\hat{\vec{\lambda}}_n$$

en donde $K_{10} = \mu_o^1$, y $\hat{\vec{\lambda}}_n = (\hat{\lambda}_o, \ldots, \hat{\lambda}_n)'$ el vector dado por

$$\begin{pmatrix} \hat{\lambda}_o \\ \hat{\lambda}_1 \\ \vdots \\ \hat{\lambda}_n \end{pmatrix} = \begin{pmatrix} 1 & \mu_o^1 & \cdots & \mu_o^n \\ \mu_o^1 & \mu_o^2 & \cdots & \mu_o^{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ \mu_o^{n+1} & \mu_o^{n+2} & \cdots & \mu_o^{2n} \end{pmatrix}^{-1} \begin{pmatrix} K_{10} \\ K_{11} \\ \vdots \\ K_{1n} \end{pmatrix}$$

"

supuesto que la función de distribución a priori $F_o(t)$ tiene más de n puntos de incremento.

El resultado se obtiene de forma totalmente análoga a la seguida en el teorema 2.4.3, utilizando ahora el teorema 1.8.7 en vez del 1.6.7.

Corolario 4.4.3. La relación entre $R_{min}(n-1)$ y $R_{min}(n)$ es

$$R_{min}(n) - R_{min}(n-1) = (\hat{\lambda}_n^*)^2 \frac{|D_n|}{|D_{n-1}|}$$

en donde $\hat{\lambda}_n^*$ es el coeficiente estimado de x^n, dentro de la clase de reglas de decisión que son polinomios de grado n, y D_n la matriz a invertir en el teorema anterior.

V. ESTIMACION DEL MOMENTO DE ORDEN i.

Ahora pretendemos calcular la regla Bayes para

$$\mu_i(F) = \int_0^\infty t^i \, dF(t) = -\int_0^\infty t^i \, dS(t) = i \int_0^\infty t^{i-1} S(t) dt = \mu_i(S)$$

en donde la primera y última igualdad se entienden como notación. Empecemos con el problema sin muestra.

Teorema 4.5.1. El estimador Bayes para $\mu_i(F)$ en el problema sin muestra es el momento de orden i respecto al origen, μ_o^i, de $F_o(t)$, la función de distribución a priori, siendo el riesgo Bayes mínimo asociado,

$$R_{min}(0) = K_{ii} - (\mu_o^i)^2$$

En efecto:

El mínimo del riesgo

$$\int_F (\mu_i(F) - d)^2 \, dP(F)$$

se alcanza para $d = E[\mu_i(F)] = \int_F \int_0^\infty t^i \, dF(t) \, d(F) = \mu_o^i$, siendo el riesgo mínimo alcanzado,

$$R_{min}(0) = V(\mu_i(F)) = E[\mu_i^2(F)] - (E[\mu_i(F)])^2 = K_{ii} - (\mu_o^i)^2$$

aplicando el teorema 1.8.7.

En cuanto al problema con muestras, vamos a considerar dos tipos de estimaciones.

<u>Estimación primera</u>:

Consideraremos una muestra de tamaño n, $T_1,\ldots,T_n$, y sea

$$\overline{\mu}_n^j = \frac{1}{n} \sum_{h=1}^{n} T_h^j, \qquad j=1,2,\ldots,i$$

el momento muestral de orden j.

Buscaremos la regla Bayes para $\mu_i(F)$ dentro de la clase de reglas de decisión de la forma

$$b_o \, \mu_o^i + b_1 \, \overline{\mu}_n^1 + b_2 \, \overline{\mu}_n^2 + \ldots + b_i \, \overline{\mu}_n^i$$

siendo μ_o^i el momento de orden i respecto al origen, de $F_o(t)$.

<u>Teorema 4.5.2</u>. En las condiciones arriba expuestas, la regla Bayes para $\mu_i(F)$ dentro de la clase de reglas de decisión de la forma

$$b_o \, \mu_o^i + b_1 \, \overline{\mu}_n^1 + b_2 \, \overline{\mu}_n^2 + \ldots + b_i \, \overline{\mu}_n^i = (\mu_o^i, \overline{\mu}_n^1, \overline{\mu}_n^2, \ldots, \overline{\mu}_n^i)\vec{b}_i$$

es

$$\hat{\mu}_i(F) = (\mu_o^i,\ \overline{\mu}_n^1,\ \overline{\mu}_n^2\ ,\ldots,\ \overline{\mu}_n^i)\vec{b}_i$$

siendo el riesgo Bayes de P

$$R_{min}(n) = K_{ii} - ((\mu_o^i)^2,\ K_{i1},\ K_{i2},\ldots,K_{ii})\hat{\vec{b}}_i$$

en donde

$$\hat{\vec{b}}_i = \begin{bmatrix} \hat{b}_o \\ \hat{b}_1 \\ \vdots \\ \hat{b}_i \end{bmatrix} = \begin{bmatrix} (\mu_o^i)^2 & \mu_o^i \mu_o^1 & \cdots & (\mu_o^i)^2 \\ \mu_o^i \mu_o^1 & \frac{1}{n}\mu_o^2 + \frac{n-1}{n} K_{11} & \cdots & \frac{1}{n}\mu_o^{i+1} + \frac{n-1}{n} K_{1i} \\ \cdots & \cdots & \cdots & \cdots \\ (\mu_o^i)^2 & \frac{1}{n}\mu_o^{i+1} + \frac{n-1}{n} K_{i1} & \cdots & \frac{1}{n}\mu_o^{2i} + \frac{n-1}{n} K_{ii} \end{bmatrix}^{-1} \begin{bmatrix} (\mu_o^i)^2 \\ K_{i1} \\ \vdots \\ K_{ii} \end{bmatrix}$$

supuesto que dicha inversa exista.

En efecto:

Siguiendo los pasos indicados en la demostración del teorema 2.5.1, al hacer mínimo el riesgo Bayes

$$R = \int_F \int_0^\infty \cdots \int_0^\infty (\mu_i(S) - \sum_{j=1}^{i} b_j\ \overline{\mu}_n^j - b_o\ \mu_o^i)^2\ dF(x_1)\ \ldots$$

$$\ldots\ dF(x_n)\ dP(F)$$

y notando que, al utilizar el teorema 1.8.7, es

$$\iint (\mu_i(S))^2 \, dF(\bar{x}) \, dP(F) = K_{ii}$$

$$\iint \bar{\mu}_n^j \, dF(\bar{x}) \, dP(F) = \mu_o^j$$

$$\iint \bar{\mu}_n^j \, \bar{\mu}_n^h \, dF(\bar{x}) \, dP(F) = \frac{1}{n} \mu_o^{h+j} + \frac{n-1}{n} K_{hj}$$

$$\iint (\bar{\mu}_n^j)^2 \, dF(\bar{x}) \, dP(F) = \frac{1}{n} \mu_o^{2j} + \frac{n-1}{n} K_{jj}$$

$$\iint \mu_i(S) \, \bar{\mu}_n^j \, dF(\bar{x}) \, dP(F) = K_{ij}$$

$$\iint \mu_i(S) \, dF(\bar{x}) \, dP(F) = \mu_o^i$$

se obtiene el resultado buscado, siendo K_{ij} los definidos en la sección anterior, y μ_o^j los momentos de $F_o(t) = 1 - S_o(t)$ en donde $S_o(t)$ es la definida en la proposición 1.8.4.

Corolario 4.5.1. Cuando $n \to \infty$, el vector de estimaciones $\hat{\vec{b}}_i = (\hat{b}_o, \hat{b}_1, \ldots, \hat{b}_i) \longrightarrow (0,0,\ldots,0,1)$ y la única información es la contenida en el momento muestral de orden i.

Corolario 4.5.2. Cuando $n \to \infty$, $R_{min}(n) \to 0$.

Vamos ahora a suponer otro tipo de estimación, utilizando una muestra de tamaño uno.

Estimación segunda:

Si x es el valor de una muestra de tamaño uno de la variable aleatoria T y si buscamos la regla Bayes dentro de la clase de re glas de decisión que sean polinomios en x, obtenemos el siguiente resultado.

Teorema 4.5.3. La regla Bayes para $\mu_i(S)$ dentro de la clase de reglas de decisión de la forma

$$\beta_o + \beta_1 x + \ldots + \beta_n x^n = (1, x, \ldots, x^n)\vec{\beta}_n$$

es

$$\hat{\mu}_i(S) = (1, x, \ldots, x^n)\hat{\vec{\beta}}_n$$

siendo el riesgo Bayes asociado

$$R_{min}(n) = K_{ii} - (\mu_o^i, K_{i1}, \ldots, K_{in})\hat{\vec{\beta}}_n$$

en donde

$$\hat{\vec{\beta}}_n = \begin{pmatrix} \hat{\beta}_o \\ \hat{\beta} \\ \vdots \\ \hat{\beta}_n \end{pmatrix} = \begin{pmatrix} 1 & \mu_o^1 & \mu_o^n \\ \mu_o^1 & \mu_o^2 & \mu_o^{n+1} \\ \cdots & \cdots & \cdots \\ \mu_o^{n+1} & \mu_o^{n+2} & \mu_o^{2n} \end{pmatrix}^{-1} \begin{pmatrix} \mu_o^i \\ K_{i1} \\ \vdots \\ K_{in} \end{pmatrix}$$

supuesto que $F_o(t)$ tiene más de n puntos de incremento.

Corolario 4.5.3. La relación entre $R_{min}(n-1)$ y $R_{min}(n)$ es

$$R_{min}(n) - R_{min}(n-1) = (\hat{\beta}_n^*)^2 \frac{|D_n|}{|D_{n-1}|}$$

en donde D_n es la matriz a invertir en el teorema anterior y $\hat{\beta}_n^*$

el coeficiente estimado para x^n, dentro de la clase de reglas de decisión que son polinomios de grado n.

VI. ESTIMACION DE LA FUNCION DE DISTRIBUCION

Trataremos ahora el problema de estimar la función de distribución F en un punto $t \geq 0$ fijo. Dado que hemos estimado ya $S(t)$ y dado que la función $g(x) = 1-x$ es lineal, utilizando el teorema 2.1.2 obtenemos dos resultados: uno para el problema sin muestra y otro para el problema con muestra.

Teorema 4.6.1. El estimador Bayes para $F(t)$ en el problema sin muestra es $1 - S_o(t) = F_o(t)$, siendo el riesgo mínimo asociado

$$R_{min}(0) = \exp\left[-\int_{[0,t)} \log\left[1+2\beta(s)(t-s)\right]d\alpha(s)\right] -$$

$$- \exp\left[-2\int_{[0,t)} \log\left[1+\beta(s)(t-s)\right]d\alpha(s)\right]$$

Teorema 4.6.2. El estimador Bayes para $F(t)$ en el problema con muestras buscada dicha regla dentro del conjunto de reglas de decisión de la forma

$$a\, F_n(t) + b\, F_o(t)$$

es

$$\hat{F}(t) = \left[1 - W_n(t)\right] F_n(t) +$$

$$+ W_n(t)\ (1 - \exp\left[-\int_{[0,t)} \log\ (1+\beta(s)(t-s))\ d\,\alpha(s)\right]$$

siendo el riesgo Bayes mínimo asociado,

$$R_{min}(n) = R_{min}(0) \cdot W_n(t)$$

en donde $W_n(t)$ es la anteriormente definida, $F_n(t) = 1 - S_n(t)$ la función de distribución muestral, y $F_o(t) = 1 - S_o(t)$ la función de distribución a priori.

Comportamiento asintótico muestral:

A partir de los teorema 1.3.10 y de los corolarios 1.3.1 y 1.3.2, o directamente a partir del teorema 4.2.3, por ser

$\hat{F}(t) = 1 - \hat{S}(t)$ obtenemos el siguiente resultado.

Teorema 4.6.3. El estimador $\hat{F}(t)$ verifica que

(a) $\hat{F}(t)$ converge casi seguro hacia $F(t)$ cuando $n \to \infty$, para cada $t \geq 0$.

(b) El riesgo Bayes mínimo alcanzado $R_{min}(n)$, converge hacia cero, cuando $n \to \infty$.

(c) $\hat{F}(t)$ converge en probabilidad hacia $F(t)$.

(d) $\hat{F}(t)$ converge en ley hacia $F(t)$.

(e) $\hat{F}(t)$ tiene una función característica que converge hacia la función característica de $F(t)$.

(f) $\hat{F}(t)$ es un estimador consistente para $F(t)$.

Tasa de convergencia del riesgo:

Dado que $W_n(t) \leq 1$ y que la tasa de convergencia es

$$\frac{R_{min}(n)}{R_{min}(n-1)} = 1 - \frac{1}{n}(1 - W_n(t))$$

la cual tiende hacia 1 al tender n hacia infinito, es

$$R_{min}(0) \geq R_{min}(n-1) \geq R_{min}(n)$$

lo cual nos expresa la conveniencia de extraer muestras, y éstas lo más grandes posibles.

VII. PROBLEMA DE LAS DOS MUESTRAS.

Sea $F(x)$ una función de distribución aleatoria, la cual supondremos es un proceso gamma extendido de parámetros $(\alpha(.), \beta(.))$. Sea $G(y)$ otra función de distribución aleatoria, la cual supondremos es un proceso gamma extendido de parámetros $(\alpha'(.), \beta'(.))$.

Lo que aquí pretendemos es el estimar, bajo pérdida cuadrática, y supuesto que $X_1,\ldots,X_m$ e $Y_1,\ldots,Y_n$ son dos muestras extraidas mediante F y G, la probabilidad

$$\Delta = P\{X \leq Y\} = \int_0^\infty F(y)dG(y).$$

Llamaremos

$$\Delta_o = \int_0^\infty F_o(y)\, dG_o(y)$$

en donde $F_o(t) = 1 - S_o(t)$ es la función de distribución a priori para $F(t)$ y $G_o(t) = 1 - S'_o(t)$ la función de distribución a priori para $G(t)$; es decir,

$$F_o(t) = 1 - S_o(t) = 1 - \exp\left[-\int_{[0,t)} \log\left[1+\beta(s)(t-s)\right] d\alpha(s)\right]$$

$$G_o(t) = 1 - S_o'(t) = 1 - \exp\left[-\int_{[0,t)} \log\left[1+\beta'(s)(t-s)\right] d\alpha'(s)\right]$$

Dado que la gundión $g(S,S') = \int_0^\infty \left[1 - S(t)\right] d\left[1 - S'(t)\right]$ es bilineal podemos aplicar el teorema 2.1.3 y obtener el siguiente resultado.

Teorema 4.7.1. La regla Bayes $\hat{\Delta}_{m,n}$ para Δ, buscada dicha regla dentro del conjunto de reglas de decisión de la forma

$$a_1 \Delta_o + a_2 \frac{1}{n} \sum_{j=1}^{n} F_o(Y_j) + a_3 \frac{1}{m} \sum_{i=1}^{m} \left[1-G_o(X_i^-)\right] + a_4 \frac{U}{n.m}$$

en donde $\frac{U}{n.m}$ es el estadístico de Mann-Whitney, es

$$\hat{\Delta}_{m,n} = W_m(t)\, W_n'(t)\Delta_o + W_m(t)\, (1 - W_n'(t))\, \frac{1}{n} \sum_{j=1}^{n} F_o(Y_j) +$$

$$+ (1 - W_m(t))\, W_n'(t)\, \frac{1}{m} \sum_{i=1}^{m} \left[1-G_o(X_i^-)\right] + (1-W_m(t))(1-W_n'(t))\, \frac{U}{n.m}$$

Obsérvese que de acuerdo con el teorema 2.1.1,
$\hat{a}_1 + \hat{a}_2 + \hat{a}_3 + \hat{a}_4 = 1.$

Comportamiento asintótico muestral:

Dado que

$$\lim_{m\to\infty} W_m(t) = 0 \qquad \text{y} \qquad \lim_{n\to\infty} W_n'(t) = 0$$

$\hat{\Delta}_{m,n}$ converge cuando $n \to \infty$ y $m \to \infty$ al estadístico de Mann-Whitney, desapareciendo toda la información a priori.

CONCLUSIONES

Como se ha comprobado, la aproximación lineal, es decir la restricción en la búsqueda de reglas Bayes a ciertas clases de reglas de decisión, que son combinaciones lineales de algún conjunto dado de funciones de la muestra, ha supuesto una gran ventaja en términos generales a la hora de las estimaciones, ya que los parámetros que en ellas intervienen tienen claras significaciones y las reglas Bayes resultan muy manejables y por tanto muy útiles, permitiéndonos en todos los casos determinar sus riesgos asociados.

Posiblemente, el resultado más importante en el primer capítulo sea la obtención de la regla Bayes

$$\hat{S}(t) = [1 - p_n(t)]\ S_n(t) + p_n(t)\ E[S(t)]$$

para la función de supervivencia $S(t)$, así como la de su riesgo mínimo asociado, verificando ésta una convergencia casi segura hacia $S(t)$. Vimos también como esa pérdida en calidad por una ganancia en utilidad, al usar $\hat{S}(t)$ en vez de la media a posteriori, era equiparable a la de usar la recta de regresión en vez de la curva de regresión.

En los capítulos segundo, tercero y cuarto continuamos el estudio del estimador $\hat{S}(t)$ del parámetro $S(t)$ cuando consideramos tres diferentes distribuciones a priori, los procesos gamma exponenciales, los procesos homogéneo simples y los procesos gamma extendidos, introduciendo entonces un parámetro c, el cual va a medir el grado de confianza en nuestra estimación a priori $S_0(t) = E[S(t)]$, de forma que cuando $c \to o$, lo cual vimos que equivalía a que nuestro $S_o(t)$ era pésimo, $\hat{S}(t)$ sólo dependerá de $S_n(t)$, la información muestral, mientras que cuando $c \to \infty$, es decir cuando nuestro conocimiento a priori era óptimo $\hat{S}(t)$, sólo va a depender de ese conocimiento a priori $S_o(t)$. Como hemos dicho, solo hemos considerado tres posibles distribuciones a priori, aunque cabría haber definido otras muchas, por ejemplo ésta nueva que a continuación damos y que llamaremos homogénea mixta que viene definida a través de su medida de Lévy asociada,

$$dN(z) = \frac{dz}{e^{cz}\ z(1 - e^{-z})}$$

y que permitiría un buen juego.

Como vemos, hemos dado una gran importancia a la estimación de la función de supervivencia y sin embargo hemos dejado un poco de lado otros parámetros. La razón fundamental nos la da los teoremas 2.1.2 y 2.1.3, los cuales vienen a decirnos que mientras transformemos el parámetro $S(t)$ mediante una función lineal, o bilineal, para convertirlo en otro parámetro, obtendremos la regla Bayes para dicho parámetro, aplicando la misma función lineal o bilineal al estimador $\hat{S}(t)$, permaneciendo los riesgos Bayes igual. Estimamos en cualquier caso la propia función de supervivencia cuando hay datos censurados, la función de distribución $F(t) = 1 - S(t)$, tratamos el problema de las dos muestras estimando $\Delta = \int_0^\infty F(y)dG(y)$ y prestamos especial interés a la estimación de la media y en general del momento de orden i, considerando diversas clases de reglas de decisión en las estimaciones.

Un camino interesante a seguir sería el estudiar que ocurre cuando los parámetros que en el estudio previo intervienen, son variables aleatorias también.

Terminamos expresando nuestra ilusión, de que el presente trabajo pueda tener una importancia práctica y confiando en que pueda servir para dar un paso más hacia adelante, aunque por supuesto se que muy pequeño, en el estudio de la decisión Bayesiana no paramétrica, teoría ésta de la que a continuación damos una amplia bibliografía.

BIBLIOGRAFIA

ANTONIAK, C.E. (1974): "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems". Ann. Statistics, vol. 2, (1.152-1.174).

BAYES, T.R. (1763): "An essay towards solving a problem in the doctrine of chances". Phil. trans. Roy. Soc. London 53, (370-418). (Reproducido en Biometrika (1958), 45, 293).

BERGER, J.O. (1980): "Statistical decision theory. Foundations, concepts, and methods". Springer-Verlag.

BERK, R.H. y SAVAGE, I.R. (1979): "Contributions to Statistics".

BLACKWELL, D. y GIRSHICK, M.A. (1979): "Theory of games and satistical decisions". Dover Publications.

BOX, G.E.P. y TIAO, G.C., (1973): "Bayesian Inference in Statistical analysis". Addishon-Wesley.

DALAL, S.R. y GAINEFORD, J.H. (1980): "On approximating parametric Bayes models by nonparametric Bayes models". Ann. Statistics, vol. 8, No. 3 (664-672).

DAVIS, M.H.A. (1977): "Linear estimation and stochastic control". Chapman and Hall.

DE FINETTI, B. (1974-75): "Theory of probability". Volúmenes I y II. John Wiley.

DE GROOT, M.H. (1970): "Optimal statistical decisions". McGraw-Hill.

DOKSUM, K. (1974): "Tailfree and neutral random probabilities and their posterior distributions". Ann. Probability, Vol. 2, No. 2, (183-201).

DOOB, J.L. (1953): "Stochastic Processes". John Wiley.

DYKSTRA, R.L. y PURUSHOTTAM LAUD (1981): "A Bayesian nonparametric approach to reliability". Ann. Statistics, Vol. 9, No. 2, (356-367).

ELANDT-JOHNSON, R.C. y JOHNSON, N.L. (1980): "Survival models and data analysis". John Wiley.

FERGUSON, T.S. (1967): "Mathematical Statistics: A decision theoretic approach". Academic Press.

FERGUSON, T.S. (1973): "A Bayesian analysis of some nonparametric problems". Ann. Statistics, Vol. 1, No. 2, (209-230).

FERGUSON, T.S. (1974): "Prior distributions on spaces of probability measures". Ann. Statistics, Vol. 2, No. 4, (615-629).

FERGUSON, T.S. y KLASS, M.J. (1972): "A representation of independent processes without Gaussian components". Ann. Math. Statistics, Vol. 43, (1634-1643).

FERGUSON, T.S. y PHADIA, E.G. (1979): "Bayesian nonparametric estimation based on censored data". Ann. Statistics, vol. 7, No. 1, (163-186).

FELLER, W. (1971): "An introduction to probability theory and its applications". Vol. 2, 2nd ed. John Wiley.

GNEDENKO, B.V. y KOLMOGOROV, A.N. (1949): "Limit distributions for sums of independent random variables", trans. K.L. CHUNG (1954). Addison-Wesley.

GOLDSTEIN, M. (1975): "Approximate Bayes solutions to some nonparametric problems". Ann. Statistics, vol. 3, No. 2, (512-517)

GOLDSTEIN, M. (1975): "A note on some Bayesian nonparametric estimates". Ann. Statistics, vol. 3, No. 3, (736-740).

GOLDSTEIN, M. (1981): "Revising previsions: A geometric interpretation". J.R. Statist. Soc. B, vol. 43, No. 2, (105-130).

HARTIGAN, J.A. (1969): "Linear Bayesian Methods". J.R. Stat. Soec., C. Vol. 3, (446-454).

HOLLANDER, M. y WOLFE, D.A. (1973): "Nonparametric statistical methods". John Wiley.

KALBFLEISCH, J.D. y PRENTICE, R.L. (1980): "The statistical analysis of failure time data". John Wiley.

KAPLAN, E.L. y MEIER, P. (1958): "Nonparametric estimation from incomplete observations". J. Am. Stat. Assoc., Vol. 53, (457-481).

KOLMOGOROV, A.N. (1933): "Foundations of the theory of probability". 2nd. ed., trans. Nathan Morrison (1956). Chelsea, New York.

LAHA, R.G. y ROHATGI, V.K. (1979): "Probability Theory". John Wiley.

LEHMANN, E.L. (1975): "Nonparametrics: Statistical methods based on ranks". Holden Day.

LEVY, P. (1937): "Théorie de l'addition des variables aléatoire". Gauthier-Villars, Paris.

LINDLEY, D.V. (1965): "Introduction to probability and statistics from a Bayesian viewpoint". Vol. 1 y 2. Cambridge University Press.

LOEVE, M. (1963): "Probability theory". 3rd. ed., Van Nostrand, Princeton, New Jersey.

MANN, H.B. y WHITNEY, D.R. (1947): "On a test of whether one of two random variables is stochastically larger than the other" Ann. Math. Statist., Vol. 18, (50-60).

PHADIA, E.G. (1980): "A note on empirical Bayes estimation of a distribution function based on censored data". Ann. Statistics, Vol. 8, No. 1, (226-299).

PHADIA, E.G. y VAN RYZIN, J. (1980): "A note on convergence rates for the product limit estimator". Ann. Statistics, Vol. 8, No. 3, (673-678).

QUESADA, V. (1972): "Métodos empírico Bayes no paramétricos". Tesis Doctoral.

RANDLES, R.H. y WOLFE, D.A. (1979): "Introduction to the theory of nonparametric statistics". John Wiley.

RIOS, S. (1967): "Métodos estadísticos". Ed. del Castillo.

RIOS, S. (1976): "Análisis de decisiones". Ed. I.C.E.

ROHATGI, V.K. (1976): "An introduction to probability theory and mathematical statistics". John Wiley.

SAVAGE, L.J. (1962): "The foundation of statistical inference". London: Methuen.

SUSARLA, V. y VAN RYZIN, J. (1976): "Nonparametric Bayesian estimation of survival curves from incomplete observations". J. Am. Statis. Associ., Vol. 71, No. 356, (897-902).

SUSARLA, V. y VAN RYZIN, J. (1980): "Large sample theory for an estimator of the mean survival time from censored samples". Ann. Statistics, Vol. 8, No. 5, (1.002-1.016).

WAGNER, S.S. y ALTMANN, S.A. (1973): "What time do the baboons come down from the trees?". Biometrics, Vol. 29 (623-635).

WILKS, S. (1962): "Mathematical Statistics". John Wiley.

WONG, E. (1971): "Stochastic processes in information and dynamical systems". Mc.Graw-Hill.

ZACKS, S. (1971): "The theory of statistical inference". John Wiley.